University of Washington
Department of Mathematics
MATH 464, Spring 2020

February 23, 2022

Homework 1

Section 1.3

Problem 1:

(a) Convert the following decimal numbers to hexadecimal form:

(i) 1023
(ii) 1025
(iii) 278.5
(iv) 14.09375
(v) 0.1240234375

Solution:

(i)

$$\begin{aligned} N &= (((a_n \cdot \alpha + a_{n-1}) \cdot \alpha + \cdots) \cdot \alpha + a_1) \cdot \alpha + a_0 \\ &= ((1 \cdot A + 0)A + 2)A + 3 \\ &= (A \cdot A + 2)A + 3 \\ &= (64 + 2)A + 3 \\ &= 3FC + 3 \\ &= 3FF \end{aligned}$$

(ii)

$$\begin{aligned} N &= (((a_n \cdot \alpha + a_{n-1}) \cdot \alpha + \cdots) \cdot \alpha + a_1) \cdot \alpha + a_0 \\ &= ((1 \cdot A + 0)A + 2)A + 5 \\ &= (A \cdot A + 2)A + 5 \\ &= 66 \cdot A + 5 \\ &= 3FC + 5 \\ &= 401 \end{aligned}$$

(iii)

$$\begin{aligned}
N_I &= (((a_n \cdot \alpha + a_{n-1}) \cdot \alpha + \cdots) \cdot \alpha + a_1) \cdot \alpha + a_0 \\
&= (2 \cdot A + 7) \cdot A + 8 \\
&= 1B \cdot A + 8 \\
&= 10E + 8 \\
&= 116 \\
y_0 &= 0.5_{10} \\
y_1 &= (16 \cdot y_0)_F = 0 \\
c_1 &= (16 \cdot y_0)_I = 8 \\
N &= 116.8_{16}
\end{aligned}$$

(iv)

$$\begin{aligned}
N_I &= 1 \cdot A + 4 \\
&= E \\
y_0 &= 0.09375 \\
y_1 &= (16 \cdot y_0)_F = 0.5 \\
c_1 &= (16 \cdot y_0)_I = 1 \\
y_2 &= (16 \cdot y_1)_F = 0 \\
c_2 &= (16 \cdot y_1)_I = 8 \\
N &= E.18
\end{aligned}$$

(v)

$$\begin{aligned}
y_0 &= 0.1240234375 \\
y_1 &= (16 \cdot y_0)_F = 0.984375 \\
c_1 &= (16 \cdot y_0)_I = 1 \\
y_2 &= (16 \cdot y_1)_F = 0.75 \\
c_2 &= (16 \cdot y_1)_I = F \\
y_3 &= (16 \cdot y_2)_F = 0 \\
c_3 &= (16 \cdot y_2)_I = C \\
N &= 0.1FC
\end{aligned}$$

(b) Convert the answers above to binary form.

(i)

$$
\begin{aligned}
c_0 &= 1023 \mod 2 = 1 \\
c_1 &= 511 \mod 2 = 1 \\
c_2 &= 255 \mod 2 = 1 \\
c_3 &= 127 \mod 2 = 1 \\
c_4 &= 63 \mod 2 = 1 \\
c_5 &= 31 \mod 2 = 1 \\
c_6 &= 15 \mod 2 = 1 \\
c_7 &= 7 \mod 2 = 1 \\
c_8 &= 3 \mod 2 = 1 \\
c_9 &= 1 \mod 2 = 1 \\
N &= 1111111111
\end{aligned}
$$

(ii)

$$
\begin{aligned}
c_0 &= 1025 \mod 2 = 1 \\
c_1 &= 512 \mod 2 = 0 \\
c_2 &= 256 \mod 2 = 0 \\
c_3 &= 128 \mod 2 = 0 \\
c_4 &= 64 \mod 2 = 0 \\
c_5 &= 32 \mod 2 = 0 \\
c_6 &= 16 \mod 2 = 0 \\
c_7 &= 8 \mod 2 = 0 \\
c_8 &= 4 \mod 2 = 0 \\
c_9 &= 2 \mod 2 = 0 \\
c_{10} &= 1 \mod 2 = 1 \\
N &= 10000000001
\end{aligned}
$$

(iii)

$$
\begin{aligned}
c_0 &= 278 \mod 2 = 0 \\
c_1 &= 139 \mod 2 = 1 \\
c_2 &= 69 \mod 2 = 1 \\
c_3 &= 34 \mod 2 = 0 \\
c_4 &= 17 \mod 2 = 1 \\
c_5 &= 8 \mod 2 = 0 \\
c_6 &= 4 \mod 2 = 0 \\
c_7 &= 2 \mod 2 = 0 \\
c_8 &= 1 \mod 2 = 1 \\
N_I &= 100010110 \\
y_0 &= 0.5_{10} \\
y_1 &= (2 \cdot 0.5)_F = 0 \\
c_1 &= (2 \cdot 0.5)_I = 1 \\
N_F &= 0.1 \\
N &= 100010110.1
\end{aligned}
$$

(iv)

$$
\begin{aligned}
c_0 &= 14 \mod 2 = 0 \\
c_1 &= 7 \mod 2 = 1 \\
c_2 &= 3 \mod 2 = 1 \\
c_3 &= 1 \mod 2 = 1 \\
N_I &= 1110 \\
y_0 &= 0.09375 \\
y_1 &= (2 \cdot 0.09375)_F = 0.1875 \\
c_1 &= (2 \cdot 0.09375)_I = 0 \\
y_2 &= (2 \cdot 0.1875)_F = 0.375 \\
c_2 &= (2 \cdot 0.1875)_I = 0 \\
y_3 &= (2 \cdot 0.375)_F = 0.75 \\
c_3 &= (2 \cdot 0.375)_I = 0 \\
y_4 &= (2 \cdot 0.75)_F = 0.5 \\
c_4 &= (2 \cdot 0.75)_I = 1 \\
y_5 &= (2 \cdot 0.5)_F = 0 \\
c_5 &= (2 \cdot 0.5)_I = 1 \\
N_F &= 0.00011 \\
N &= 1110.00011
\end{aligned}
$$

(v)

$$
\begin{aligned}
y_0 &= 0.1240234375 \\
y_1 &= (2 \cdot 0.1240234375)_F = 0.248046875 \\
c_1 &= (2 \cdot 0.1240234375)_I = 0 \\
y_2 &= (2 \cdot 0.248046875)_F = 0.49609375 \\
c_2 &= (2 \cdot 0.248046875)_I = 0 \\
y_3 &= (2 \cdot 0.49609375)_F = 0.9921875 \\
c_3 &= (2 \cdot 0.49609375)_I = 0 \\
y_4 &= (2 \cdot 0.9921875)_F = 0.984375 \\
c_4 &= (2 \cdot 0.9921875)_I = 1 \\
y_5 &= (2 \cdot 0.984375)_F = 0.96875 \\
c_5 &= (2 \cdot 0.984375)_I = 1 \\
y_6 &= (2 \cdot 0.96875)_F = 0.9375 \\
c_6 &= (2 \cdot 0.96875)_I = 1 \\
y_7 &= (2 \cdot 0.9375)_F = 0.875 \\
c_7 &= (2 \cdot 0.9375)_I = 1 \\
y_8 &= (2 \cdot 0.875)_F = 0.75 \\
c_8 &= (2 \cdot 0.875)_I = 1 \\
y_9 &= (2 \cdot 0.75)_F = 0.5 \\
c_9 &= (2 \cdot 0.75)_I = 1 \\
y_{10} &= (2 \cdot 0.5)_F = 0 \\
c_{10} &= (2 \cdot 0.5)_I = 1 \\
N &= 0.0001111111
\end{aligned}
$$

(c) Prove that any proper fraction that has a terminating hexadecimal expansion also has a terminating decimal expansion. Explain in general terms why the converse of this statement is not valid.

Solution:

Since a fraction only has a terminal expansion in a base β where the prime factors of the denominator is also a prime factors of β, a terminating hexadecimal expansion with 2 as the only prime factor, also has a terminating decimal expansion, which has $2, 5$ as prime factors. The converse is not true because 5 is not a prime factor of base 16. An example is $0.1_{10} = 0.19999\cdots_{16}$

(d) For the computer example that computed the sums of reciprocals, compute the relative errors for the operations corresponding to k = 1000, 1006, 1012, 1018.

Solution:

For $k = 1000$, absolute error and relative error $= 0.0000122$. For $k = 1006$, absolute error and relative error $= 0.0000088$. For $k = 1012$, absolute error and relative error $= 0.0000157$. For $k = 1018$, absolute error and relative error $= 0.0000322$.

(e) (i) If x is a real number and $\bar{x}$ is its chopped machine representation on a computer with base 10 and mantissa length m, show that the relative error, $\frac{|x-\bar{x}|}{|x|}$is bounded by 10^{-m+1}.

To show $\frac{|x-\bar{x}|}{|x|}$is bounded by 10^{-m+1}, we have to show that $|x - \bar{x}|$ is bounded above by something, and $|x|$ is bounded below by something.

$$\begin{aligned} \bar{x} &= (0.d_1d_2d_3 \cdots d_m)_\beta \beta^C \\ |x - \bar{x}| &\leq \beta^{\mu-m} \\ |x| &=\geq \beta^{\mu-1} \\ \frac{|x - \bar{x}|}{|x|} &\leq \beta^{-m+1} \\ \beta &= 10 \\ \frac{|x - \bar{x}|}{|x|} &\leq 10^{-m+1} \end{aligned}$$

(ii) If $\delta = 10^{-m+1}$ (in the case of chopping) or $\delta = 0.5x10^{-m+1}$ (in the case of symmetric rounding), show that $\bar{x} = x(1 + \epsilon)$ where $|\epsilon| \leq \delta$.
We can rearrange the function

$$\begin{aligned} \bar{x} &= x(1 + \epsilon) \\ \bar{x} &= x + x\epsilon \\ \frac{\bar{x} - x}{x} &= \epsilon \end{aligned}$$

We have shown that $\frac{|\bar{x}-x|}{|x} | \leq \beta^{-m+1}$ in case of chopping, and analogously we can obtain $\frac{|\bar{x}-x|}{|x} | \leq \frac{1}{2}\beta^{-m+1}$ in case of symmetric rounding. Hence, for $\beta = 10$, there necessarily exists an ϵ such that $-10^{-m+1} \leq \epsilon \leq 10^{-m+1}$ for chopping case, and $-0.5 \times 10^{-m+1} \leq \epsilon \leq 0.5 \times 10^{-m+1}$. Therefore, for both the chopping and symmetric round case, there exists ϵ such that $\bar{x} = x(1 + \epsilon)$ where $|\epsilon| \leq \delta$

(iii) Given a function $F(x), F'$ continuous, we have from the mean-value theorem that in evaluating F at x the relative error from this source alone is

$$\frac{F(\bar{x}) - F(x)}{F(x)} = \frac{F(x + \epsilon x) - F(x)}{F(x)} = \frac{F'(e)[(x + \epsilon x) - x]}{F(x)} = \epsilon x \frac{F'(x)}{F(x)}.$$

Analogously, $F(\bar{x}) - F(x) \simeq \epsilon x F'(x)$. Use these formulas to give estimates of absolute and relative error if

(i) $F(x) = x^k$, for various values of k and x;
(ii) $F(x) = e^x$, for large values of x;
(iii) $F(x) = \sin(x)$, for small values of x;
(iv) $F(x) = x^2 - 1$, for x near 1;
(v) $F(x) = \cos(x)$, for x near $\frac{\pi}{2}$.

Solution:

(i)

$$\begin{aligned} F(\bar{x}) - F(x) &\simeq \epsilon x F'(x) \\ &\simeq \epsilon k x^k \\ \frac{F(\bar{x}) - F(x)}{F(x)} &\simeq \epsilon x \frac{F'(x)}{F(x)} \\ &\simeq \epsilon k \end{aligned}$$

(ii)

$$\begin{aligned} F(\bar{x}) - F(x) &\simeq \epsilon x F'(x) \\ &\simeq \epsilon x e^x \\ \frac{F(\bar{x}) - F(x)}{F(x)} &\simeq \epsilon x \frac{F'(x)}{F(x)} \\ &\simeq \epsilon x \end{aligned}$$

(iii)

$$\begin{aligned} F(\bar{x}) - F(x) &\simeq \epsilon x F'(x) \\ &\simeq \epsilon x \cos(x) \\ \frac{F(\bar{x}) - F(x)}{F(x)} &\simeq \epsilon x \frac{F'(x)}{F(x)} \\ &\simeq \epsilon x \frac{\cos(x)}{\sin(x)} \\ \lim_{x \to 0} \epsilon x \frac{\cos(x)}{\sin(x)} &= \lim_{x \to 0} \frac{\epsilon \cos(x) - \epsilon x \sin(x)}{\cos(x)} \\ &= \epsilon \end{aligned}$$

(iv)

$$\begin{aligned} F(\bar{x}) - F(x) &\simeq \epsilon x F'(x) \\ &\simeq \epsilon 2x \\ \frac{F(\bar{x}) - F(x)}{F(x)} &\simeq \epsilon x \frac{F'(x)}{F(x)} \\ &\simeq \epsilon \frac{2x}{x^2 - 1} \\ \lim_{x \to 1} \epsilon \frac{2x}{x^2 - 1} &= \lim_{x \to 1} \frac{4\epsilon x}{2x} \\ &= 2\epsilon \end{aligned}$$

(v)

$$\begin{aligned} F(\bar{x}) - F(x) &\simeq \epsilon x F'(x) \\ &\simeq -\epsilon x \sin(x) \\ \frac{F(\bar{x}) - F(x)}{F(x)} &\simeq \epsilon x \frac{F'(x)}{F(x)} \\ &\simeq -\epsilon x \frac{\sin(x)}{\cos(x)} \\ \lim_{x \to \frac{\pi}{2}} -\epsilon x \frac{\sin(x)}{\cos(x)} &= \lim_{x \to \frac{\pi}{2}} \qquad\qquad = \epsilon \end{aligned}$$

Problem 8:

Consider a computer with mantissa length $m = 3$, base $b = 10$, and exponent constraints $\mu = -2 \leq c \leq 2 = M$. How many real numbers will this computer represent exactly?

Solution:

Given the mantissa has length $m = 3$, assume that the representation format is normalized floating point number, the first digit has 9 choices (cannot be 0), the second digit has 10 choices, and the third digit has 10 choices. Hence, the total number of representation only considering the mantissa is $9 \cdot 10 \cdot 10 = 900$ numbers. However, as we know the exponent is $2 \leq c \leq 2$, we can have the radix point at 5 different places. We can thus represent $5 \cdot 900 = 4500$ numbers, considering the exponent constraints. We also have negatives, which means we can represent $2 \cdot 4500 = 9000$ numbers. We can finally add the exception 0 to our set of numbers, which gives us 9001 numbers.

Section 2.1

Problem 3:

How many multiplications must be performed to form the product of two $n \times n$ matrices?

Solution:

For $n \times n$ matrix A, B, we need n^2 multiplications to calculate Ab_i where b_i is a column of B. Hence to calculate AB we need $n^2 \cdot n = n^3$ multiplications.

Problem 4:

How many multiplications must be performed to form the product $L\mathbf{x}$ when L is an $n \times n$ lower triangular matrix and $\mathbf{x}$ is an $n \times 1$ column vector? (Do not count multiplications by zero.)

Solution:

$$L = \begin{bmatrix} L_{11} & 0 & 0 & \cdots & 0 \\ L_{21} & L_{22} & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & 0 \\ L_{(n-1)1} & L_{((n-1)2} & \cdots & L_{(n-1)(n-1)} & 0 \\ L_{n1} & L_{n2} & L_{n3} & \cdots & L_{nn} \end{bmatrix}, \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{n-1} \\ x_n \end{bmatrix}$$

$L\mathbf{x}$ has 1 multiplication in row 1, 2 multiplications in row 2, and inductively we have n multiplications in row n. Therefore the total number of multiplication is

$$\sum_{k=1}^{n} k = \frac{n(n+1)}{2}$$

Problem 6:

how many multiplications must be performed to evaluate the determinant of an $(n \times n)$ matrix according to the procedure of Problem 5 when $n = 3$, when $n = 4$, when $n = 5$, and for arbitrary n? (For $n = 10$, more than 3,000,000 multiplications are required. If a person were able to multiply two numbers and record the result at a rate of one per second, it would require 126 eight-hour days to find the determinant of a (10×10) matrix using a cofactor expansion.)

Solution:

$$A \in \mathbb{R}^{n \times n}$$

$$\text{mult}(\det(A)) = n(\cdot\text{mult}(\det(B)) + 1), B \in \mathbb{R}^{(n-1)\times(n-1)}$$

When $n = 3$, there are 9 multiplications. When $n = 4$, there are 40 multiplications. When $n = 5$, there are 205 multiplications.

Problem 10:

Let T and S be $(n \times n)$ upper triangular matrices. Use the definition of matrix multiplication to show that the product ST is also upper triangular.

Solution:

$$S = \begin{bmatrix} s_{11} & s_{12} & s_{13} & \cdots & s_{1n} \\ 0 & s_{22} & s_{23} & \cdots & s_{2n} \\ 0 & 0 & s_{33} & \cdots & s_{3n} \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & s_{nn} \end{bmatrix}$$

$$T = \begin{bmatrix} t_{11} & t_{12} & t_{13} & \cdots & t_{1n} \\ 0 & t_{22} & t_{23} & \cdots & t_{2n} \\ 0 & 0 & t_{33} & \cdots & t_{3n} \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & t_{nn} \end{bmatrix} \quad ST = \begin{bmatrix} s_{11}t_{11} & s_{11}t_{12} + s_{12}t_{22} & s_{11}t_{13} + s_{12}t_{23} + s_{13}t_{33} & \cdots \\ 0 & s_{22}t_{22} & s_{22}t_{23} + s_{23}t33 & \cdots \\ 0 & 0 & s_{44}t_{44} & \cdots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & s_{nn}t_{nn} \end{bmatrix}$$

Problem 11:

Let A be a lower(upper) triangular nonsingular matrix. Show that A^{-1} is also lower (upper) triangular. [Hint: Consider the equation $AA^{-1} = I$, entry by entry.] Let A be a lower triangular nonsingular matrix.

$$A = \begin{bmatrix} a_{11} & 0 & \cdots & 0 & 0 & 0 \\ a_{21} & a_{22} & 0 & \cdots & 0 & 0 \\ a_{31} & a_{32} & a_{33} & 0 & \ddots & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\ a_{(n-1)1} & a_{(n-1)2} & a_{(n-1)3} & \cdots & a_{(n-1)(n-1)} & 0 \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{n(n-1)} & a_{nn} \end{bmatrix}$$

We know $AA^{-1} = I$. Let $B = A^{-1}$. Consider AA^{-1} from the last column of B and I.

$$AB = I$$

$$\begin{bmatrix} a_{11} & 0 & \cdots & 0 & 0 & 0 \\ a_{21} & a_{22} & 0 & \cdots & 0 & 0 \\ a_{31} & a_{32} & a_{33} & 0 & \ddots & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\ a_{(n-1)1} & a_{(n-1)2} & a_{(n-1)3} & \cdots & a_{(n-1)(n-1)} & 0 \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{n(n-1)} & a_{nn} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1(n-1)} & b_{1n} & \\ b_{21} & b_{22} & \cdots & b_{2(n-1)} & b_{2n} & \\ b_{31} & b_{32} & \cdots & b_{3(n-1)} & b_{3n} & \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ b_{(n-1)1} & b_{(n-1)2} & \cdots & b_{(n-1)(n-1)} & b_{(n-1)n} & \\ b_{n1} & b_{n2} & \cdots & b_{n(n-1)} & b_{nn} & \end{bmatrix} = I$$

Let $\mathbf{x}$ be the nth column of B. It is clear that

$$Ax = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1(n-1)}x_{n-1} + a_{1n}x_n \\ a_{11}x_1 + a_{12}x_2 + \cdots + a_{1(n-1)}x_{n-1} + a_{1n}x_n \\ \vdots \\ a_{(n-1)1}x_1 + a_{(n-1)2}x_2 + \cdots + a_{(n-1)(n-1)}x_{n-1} + a_{(n-1)n}x_n \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{n(n-1)}x_{n-1} + a_{nn}x_n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}$$

Since A is lower triangular nonsingular, we first notice that the first row of $Ax = I_n$ is just $a_{11}x_1 = 0$. Since $a_{11} \neq 0$, we have $x_1 = 0$. Substitute $x_1 = 0$ back to $Ax = I$, by the same analogy, we can find $x_2 = 0$. If we keep doing forward substitution and solving for x_i, we soon discover that

$$x = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ x_n \end{bmatrix}$$

where x is the nth column of $B = A^{-1}$. We apply the method inductively backwards from $(n-1)$th column of B to the first column of B, column by column, we will find B necessarily a lower triangular. By $B = A^{-1}$, A^{-1} is lower triangular.

Let A be upper triangular nonsingular matrix. We know $AA^{-1} = I$. Let $B = A^{-1}$. Consider AA^{-1} starting from the first column of B and I.

$$AB = I$$

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1(n-1)} & a_{1n} \\ 0 & a_{22} & a_{23)} & \cdots & a_{2(n-1)} & a_{2n} \\ 0 & 0 & a_{33} & \cdots & a_{3(n-1)} & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & a_{(n-1)(n-1)} & a_{(n-1)n} \\ 0 & 0 & \cdots & 0 & 0 & a_{nn} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1(n-1)} & b_{1n} & \\ b_{21} & b_{22} & \cdots & b_{2(n-1)} & b_{2n} & \\ b_{31} & b_{32} & \cdots & b_{3(n-1)} & b_{3n} & \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ b_{(n-1)1} & b_{(n-1)2} & \cdots & b_{(n-1)(n-1)} & b_{(n-1)n} & \\ b_{n1} & b_{n2} & \cdots & b_{n(n-1)} & b_{nn} & \end{bmatrix} = I$$

Let x be the first column of B. This is similar to the lower triangular nonsingular case. The only difference is we will find $x_n = 0$, and we use backward substitution to find $x_(n-1), x_(n-2), \cdots, x_1$. For column vector x, aside from x_1 can be non-zero, x_2 to x_n are all 0. We apply this method inductively to the 2nd to nth column of B, column by column, we will find B also necessarily upper triangular. By $B = A^{-1}$, A^{-1} is also upper triangular.

University of Washington
Department of Mathematics
MATH 464, Spring 2020

February 23, 2022

Homework 2

Section 2.2.4

Problem 4:

In Example 2.8 replace .0001 by 10^{-n}. Use three-digit floating-decimal calculations for solving the system without pivoting, and determine the positive integer values of n for which the computed solution is "significantly" different from the true solution.

Solution:

Replacing .0001 by 10^{-n} in Example 2.8 gives us the following:

$$\begin{aligned} 10^{-n}x + 1.00y &= 1.00 \\ 1.00x + 1.00y &= 2.00 \end{aligned}$$

Applying Gaussian Elimination we get

$$\begin{aligned} 10^{-n}x + 1.00y &= 1.00 \\ \left(1 - \frac{1}{10^{-n}}\right) y &= 2 - \frac{1}{10^{-n}} \end{aligned}$$

Simplify and we obtain

$$\begin{aligned} x &= \frac{-10^n}{1 - 10^n} \\ y &= \frac{2 - 10^n}{1 - 10^n} \end{aligned}$$

For $n = 3$, the true solution is

$$\begin{aligned} x &= \frac{1000}{999} \approx 1.00100 \\ y &= \frac{998}{999} \approx 0.99990 \end{aligned}$$

and the computed solution is

$$\begin{aligned} x &= 1.00 \\ y &= 0.999 \end{aligned}$$

For $n = 4$, the true solution is

$$x = \frac{10000}{9999} \approx 1.00010$$
$$y = \frac{9998}{9999} \approx 0.99990$$

and the computed solution is

$$x = 0.00$$
$$y = 1.00$$

For $n \geq 4$, computations on this machine are unreliable, since the computed solution differs significantly from the true solution.

Problem 5:
Repeat Problem 4 for the system

$$10^{-n}x_1 + x_2 = 3$$
$$x_1 - x_2 = -2.$$

Solution:
Applying Gaussian Elimination we get

$$10^{-n}x_1 + x_2 = 3$$
$$\left(-1 + -\frac{1}{10^{-n}}\right) x_2 = -2 + -\frac{3}{10^{-n}}$$

Simply and we get

$$x_1 = \frac{10^n}{1 + 10^n}$$
$$x_2 = \frac{2 + 3 \cdot 10^n}{1 + 10^n}$$

For $n = 2$, the true solution is

$$x_1 = \frac{100}{101} \approx 0.99010$$
$$x_2 = \frac{302}{101} \approx 2.99010$$

and the computed solution is

$$x_1 = 1.00$$
$$x_2 = 2.99$$

For $n = 3$, the true solution is

$$x_1 = \frac{1000}{1001} \approx 0.99900$$
$$x_2 = \frac{3002}{1001} \approx 2.99900$$

and the computed solution is

$$x_1 = 0.00$$
$$x_2 = 3.00$$

For $n \geq 3$, computations on this machine are unreliable, since the computed solution differs significantly from the true solution.

Problem 9:

Given the system of equations (2.6), show that $\frac{n(n+1)}{2}$ multiplications are required to solve for $x_n, x_{n-1}, \ldots, x_2, x_1$.

Solution:

System of equations (2.6) is the following:

$$\begin{aligned}
a'_{11}x_1 + a'_{12}x_2 + a'_{13}x_3 + \cdots + a'_{1,n-1}x_{n-1} + a'_{1n}x_n &= b'_1 \\
a'_{22}x_2 + a'_{23}x_3 + \cdots + a'_{2,n-1}x_{n-1} + a'_{2n}x_n &= b'_2 \\
a'_{33}x_3 + \cdots + a'_{3,n-1}x_{n-1} + a'_{3n}x_n &= b'_3 \\
&\vdots \\
a'_{n-1,n-1}x_{n-1} + a'_{n-1,n}x_n &= b'_{n-1} \\
a'_{n,n}x_n &= b'_n
\end{aligned}$$

We need to see the number of multiplications are required to solve for $x_n, x_{n-1}, \ldots, x_2, x_1$ by backsolving. We notice that solving the last equation requires 1 multiplication to find x_n. For x_{n-1}, we need to multiply x_n with $a_{n-1,n}$, and then multiply by $\frac{1}{a'_{n-1,n-1}}$ to find x_{n-1}. Solving for x_{n-2} follows suit, where we have to plug in the solved x_n, x_{n-1}. For solving x_{k-1}, we have to plug in $x_n, x_{n-1}, \ldots, x_{k+1}, x_k$, which is $n - k + 1$ multiplications to make the equation at row $k - 1$ solvable with 1 more multiplication. Hence, to account for the multiplications at each row, we have the following equation:
Number of multiplications required to solve for $x_n, x_{n-1}, \ldots, x_2, x_1$ is equal to

$$\begin{aligned}
\sum_{k=1}^{n}(n - k + 1) &= n + (n-1) + (n-2) + \cdots + 2 + 1 \\
&= \frac{n(n+1)}{2}
\end{aligned}$$

Problem 10:

Construct a (3×3) coefficient matrix where the implicit-scaling row interchanges are different from the row interchanges of partial pivoting.

Solution:

$$A = \begin{bmatrix} 1 & 2 & 6 \\ 3 & 4 & 5 \\ 2 & 3 & 4 \end{bmatrix}, d = \begin{bmatrix} 6 \\ 5 \\ 4 \end{bmatrix}$$

For partial pivoting, we choose the smallest $I \geq k$ for which $\left|a_{Ik}^{(k-1)}\right|$ is the maximum of

$$\left|a_{kk}^{(k-1)}\right|, \left|a_{k+1,k}^{(k-1)}\right|, \ldots, \left|a_{nk}^{(k-1)}\right|.$$

For coefficient matrix A, we switch row 1 and 2. For implicit-scaling, we choose the pivot near the smallest $I \geq k$ for which

$$\frac{\left|a_{Ik}^{(k-1)}\right|}{d_I}$$

is the maximum of

$$\frac{\left|a_{Ik}^{(k-1)}\right|}{d_k}, \ldots, \frac{\left|a_{nk}^{(k-1)}\right|}{d_n}.$$

For coefficient matrix A, we switch row 1 and 3.

Section 2.3

Problem 3:

Let $\mathbf{x}$ be any vector in R^n as given by (2.36), and let $\{\mathbf{x}^{(j)}\}_{j=1}^{\infty}$ be any sequence of vectors in R^n.

(a) Show that $\|\mathbf{x}\|_1 \geq \|\mathbf{x}\|_2 \geq \|\mathbf{x}\|_\infty \geq \frac{1}{n}\|\mathbf{x}\|_1$ for all $\mathbf{x} \in R^n$.

(b) From the definition of convergence of a sequence of real numbers, use Part(a) to show that if $\lim_{j\to\infty} \|\mathbf{x}^{(j)}\|_p = 0$ for $p = 1, 2$, or ∞, then $\lim_{j\to\infty} \|\mathbf{x}^{(j)}\|_q = 0$ for $q = 1, 2$, or $\infty, q \neq p$.

(c) For p given as $1, 2$, or ∞, show that $\lim_{j\to\infty} \|\mathbf{x}^{(j)} - \mathbf{x}\|_p = 0$ implies that

$$\lim_{j\to\infty} \|A(\mathbf{x}^{(j)} - \mathbf{x})\|_p \equiv \lim_{j\to\infty} \|A\mathbf{x}^{(j)} - A\mathbf{x}\|_p = 0.$$

Solution:

(a) First, $\|\mathbf{x}\|_1 = |x_1| + |x_2| + \cdots + |x_n|$. By the Multinomial Theorem, we can set $n = 2$, and obtain the following:

$$\begin{aligned} \|\mathbf{x}\|_1^2 &= |x_1|^2 + |x_2|^2 + \cdots + |x_n|^2 + |x_1||x_2| + |x_1||x_3| + \cdots \\ &= \|\mathbf{x}\|_2^2 + |x_1||x_2| + |x_1||x_3| + \cdots \\ &\geq \|\mathbf{x}\|_2^2 \end{aligned}$$

which proves the first part of the inequality. Moving on, since $\|\mathbf{x}\|_\infty = \max(|x_1|, \cdots, |x_n|)$, it is clear that

$$\begin{aligned}\|\mathbf{x}\|_2^2 &= |x_1|^2 + \cdots + |x_n|^2 \\ &\geq (\max(|x_1|, \cdots, |x_n|))^2 \\ &\geq \|\mathbf{x}\|_\infty^2\end{aligned}$$

To show that $\mathbf{x}_\infty \geq \frac{1}{n}\|\mathbf{x}\|_1$, we start from the definition of the $\|\mathbf{x}\|_\infty$:

$$\begin{aligned}\|\mathbf{x}\|_\infty &= \max(|x_1|, \cdots, |x_n|) \\ &\geq |x| \in (|x_1|, \cdots, |x_n|) \\ n \cdot \|\mathbf{x}\|_\infty &\geq \sum_{k=1}^{n} |x_k| \\ &\geq \|\mathbf{x}\|_1 \\ \|\mathbf{x}\|_\infty &\geq \frac{1}{n}\|\mathbf{x}\|_1\end{aligned}$$

(b) The statement in (a) guarantees that for $p, q = 1, 2, \infty$, regardless of which p, q is chosen, we have $\|\mathbf{x}\|_p \geq \frac{1}{n}\|\mathbf{x}\|_q$. Since we have $\lim_{j\to\infty} \|\mathbf{x}^{(j)}\|_p = 0$ for $p = 1, 2, \infty$, we have

$$\begin{aligned}n \lim_{j\to\infty} \|\mathbf{x}^{(j)}\|_p &\geq \lim_{j\to\infty} \|\mathbf{x}^{(j)}\|_q \\ \lim_{j\to\infty} \|\mathbf{x}^{(j)}\|_q &\geq \frac{1}{n} \lim_{j\to\infty} \|\mathbf{x}^{(j)}\|_p\end{aligned}$$

Both $n \lim_{j\to\infty} \|\mathbf{x}^{(j)}\|_p = \frac{1}{n} \lim_{j\to\infty} \|\mathbf{x}^{(j)}\|_p = 0$. By the squeeze theorem,

$$n \lim_{j\to\infty} \|\mathbf{x}^{(j)}\|_p = 0 \leq \lim_{j\to\infty} \|\mathbf{x}^{(j)}\|_q \leq \frac{1}{n} \lim_{j\to\infty} \|\mathbf{x}^{(j)}\|_p = 0$$

Therefore $\lim_{j\to\infty} \|\mathbf{x}^{(j)}\|_q = 0$ for $q = 1, 2$, or $\infty, q \neq p$

(c) We know from the properties of operator norm that for all $\mathbf{x}$

$$\|A\mathbf{x}\| \leq \|A\|\|\mathbf{x}\|$$

Since $\lim_{j\to\infty} \|\mathbf{x}^{(j)} - \mathbf{x}\|_p = 0$,

$$\begin{aligned}\lim_{j\to\infty} \|A\|_p \|\mathbf{x}^{(j)} - \mathbf{x}\|_p &= 0 \\ \lim_{j\to\infty} \|A\|_p \|\mathbf{x}^{(j)} - \mathbf{x}\|_p &\geq \lim_{j\to\infty} \|A(\mathbf{x}^{(j)} - \mathbf{x})\|_p\end{aligned}$$

Problem 7:

Let A be the (4×4) coefficient matrix of the system in Example 2.6. Then A^{-1} is given by

$$A^{-1} = \begin{bmatrix} 68 & -41 & -17 & 10 \\ -41 & 25 & 10 & -6 \\ -17 & 10 & 5 & -3 \\ 10 & -6 & -3 & 2 \end{bmatrix}$$

Find the condition number $\|A\|_\infty \|A^{-1}\|_\infty$ of this matrix. Use the inequality (2.45) of Theorem 2.1 and the computer results of Example 2.7 to estimate the relative error $\frac{\|\mathbf{x}_c - \mathbf{x}_t\|_\infty}{\|\mathbf{x}_t\|_\infty}$. What is the true relative error?

Solution:

$$A = \begin{bmatrix} 5 & 7 & 6 & 5 \\ 7 & 10 & 8 & 7 \\ 6 & 8 & 10 & 9 \\ 5 & 7 & 9 & 10 \end{bmatrix}$$

$$\|A\|_\infty = 33, \|A^{-1}\|_\infty = 136$$

The condition number is 4488.

$$\frac{1}{\|A\|_\infty \|A^{-1}\|_\infty} \frac{\|A\mathbf{x}_c - \mathbf{b}\|_\infty}{\|\mathbf{b}\|_\infty} \leq \frac{\|\mathbf{x}_c - \mathbf{x}_t\|_\infty}{\|\mathbf{x}_t\|_\infty} \leq \|A\|_\infty \|A^{-1}\|_\infty \frac{\|A\mathbf{x}_c - \mathbf{b}\|_\infty}{\|\mathbf{b}\|_\infty}$$

$$\frac{1}{4488} \frac{0.15 \cdot 10^{-4}}{33} \leq \frac{\|\mathbf{x}_c - \mathbf{x}_t\|_\infty}{\|\mathbf{x}_t\|_\infty} \leq 4488 \frac{0.15 \cdot 10^{-4}}{33}$$

$$0.10128 \cdot 10^{-9} \leq \frac{\|\mathbf{x}_c - \mathbf{x}_t\|_\infty}{\|\mathbf{x}_t\|_\infty} \leq 0.204 \cdot 10^{-2}$$

Problem 8:

Let A_n be the (2×2) matrix given by

$$A_n = \begin{bmatrix} 1 & 2 \\ 2 & 4 + \frac{1}{n^2} \end{bmatrix}.$$

Find A_n^{-1} and the condition number $\|A_n\|_\infty \|A_n^{-1}\|_\infty$. Let $n = 100$ so that

$$A_n = \begin{bmatrix} 1 & 2 \\ 2 & 4.0001 \end{bmatrix},$$

and let

$$\mathbf{b} = \begin{bmatrix} 1 \\ 2 - \frac{1}{n^2} \end{bmatrix} = \begin{bmatrix} 1 \\ 1.9999 \end{bmatrix}.$$

Solve $A_{100}\mathbf{x} = \mathbf{b}$ mathematically and called the answer $\mathbf{x}_t$. Let $\mathbf{x}_c$ be given by

$$\mathbf{x}_c = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

Find $\mathbf{r} = A_{100}\mathbf{x}_c - \mathbf{b}$ and check the error bound $\|\mathbf{x}_c - \mathbf{x}_t\|_\infty \leq \|A^{-1}\|_\infty \|\mathbf{r}\|_\infty$. One might expect from this problem and from Example 2.14 that ill-conditioned matrices must have small determinants. Find the determinant of the matrix A in Problem 7 to see that this suspicion is not always valid.

Solution:

$$A_n^{-1} = \begin{bmatrix} 1+4n^2 & -2n^2 \\ -2n^2 & n^2 \end{bmatrix}.$$

The condition number is $(6+\frac{1}{n^2})(1+6n^2)$.

$$\begin{aligned} A_{100}\mathbf{x} &= \mathbf{b} \\ \mathbf{x}_t &= \begin{bmatrix} 1+40000 & -20000 \\ -20000 & 10000 \end{bmatrix} \begin{bmatrix} 1 \\ 1.9999 \end{bmatrix} \\ &= \begin{bmatrix} 3 \\ -1 \end{bmatrix} \\ \mathbf{r} &= \begin{bmatrix} 1 & 2 \\ 2 & 4.0001 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} - \begin{bmatrix} 3 \\ -1 \end{bmatrix} \\ &= \begin{bmatrix} 1 \\ 2 \end{bmatrix} - \begin{bmatrix} 3 \\ -1 \end{bmatrix} \\ &= \begin{bmatrix} -2 \\ 3 \end{bmatrix} \\ \|\mathbf{x}_c - \mathbf{x}_t\|_\infty &= 2 \\ \|A^{-1}\|_\infty \|\mathbf{r}\|_\infty &= 60001 \cdot 3 = 180003 \\ 2 = \|\mathbf{x}_c - \mathbf{x}_t\|_\infty &\leq \|A^{-1}\|_\infty \|\mathbf{r}\|_\infty = 180003 \end{aligned}$$

To find the determinant for Problem 7 we first do Gaussian Elimination

$$\begin{aligned} A &= \begin{bmatrix} 5 & 7 & 6 & 5 \\ 7 & 10 & 8 & 7 \\ 6 & 8 & 10 & 9 \\ 5 & 7 & 9 & 10 \end{bmatrix} \\ &= \begin{bmatrix} 5 & 7 & 6 & 5 \\ 0 & -\frac{1}{5} & -\frac{2}{5} & 0 \\ 0 & 0 & 2 & 3 \\ 0 & 0 & 0 & \frac{1}{2} \end{bmatrix} \end{aligned}$$

The determinant of A is the diagonal product of the reduced matrix $5(-\frac{1}{5}) \cdot 2(\frac{1}{2}) = -1$.

Problem 9:

For the matrix A_n in Problem 8, find a (2×2) singular matrix B as in (2.46) such that

$$\left(\frac{\|A_n - B\|}{\|A_n\|} - \frac{1}{\kappa(A_n)} \right) < \frac{1}{n^2}$$

$$\left(\frac{\|A_n - B\|}{\|A_n\|} - \frac{1}{\kappa(A_n)} \right) = \left(\frac{\|A_n - B\|}{6+\frac{1}{n^2}} - \frac{1}{(6+\frac{1}{n^2})(1+6n^2)} \right)$$

Let $B = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$, we have

$$\begin{aligned} \frac{\frac{1}{n^2}}{6+\frac{1}{n^2}} - \frac{1}{(6+\frac{1}{n^2})(1+6n^2)} &= \frac{1}{n^2} \cdot \frac{1}{6+\frac{1}{n^2}} - \frac{1}{(6+\frac{1}{n^2})(1+6n^2)} \\ &= \frac{\frac{1+6n^2}{n^2} - 1}{(6+\frac{1}{n^2})(1+6n^2)} \\ &= \frac{1+5n^2}{(1+6n^2)^2} \end{aligned}$$

and we can see clearly

$$\frac{1}{n^2} \cdot \frac{1}{6+\frac{1}{n^2}} - \frac{1}{(6+\frac{1}{n^2})(1+6n^2)} < \frac{1}{n^2}$$

Problem 18:

If an $(n \times n)$ matrix A has an LU-decomposition where L and U are known, what are the two necessary steps for solving the system $A^T\mathbf{y} = \mathbf{b}$? Find an LU-decomposition of the (3×3) coefficient matrix in Example 2.2, Section 2.1, and use the decomposition in this manner to solve $A^T\mathbf{y} = \mathbf{e_1}$.

Solution:

$$\begin{aligned} A^T &= (LU)^T \\ &= U^T L^T \\ U^T L^T \mathbf{y} &= \mathbf{b} \end{aligned}$$

Now, U^T is lower triangular, and L^T is upper triangular. We first solve for

$$U^T \mathbf{z} = \mathbf{b}$$

then

$$L^T \mathbf{y} = \mathbf{z}$$

For Example 2.2, Section 2.1

$$A = \begin{bmatrix} 1 & 0 & 1 \\ -1 & 1 & -1 \\ 2 & 1 & 0 \end{bmatrix}$$

$$L = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 2 & 1 & 1 \end{bmatrix}$$

$$U = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{bmatrix}$$

$$U^T\mathbf{z} = \mathbf{e_1}$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & -2 \end{bmatrix} \mathbf{z} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

$$\mathbf{z} = \begin{bmatrix} 1 \\ 0 \\ \frac{1}{2} \end{bmatrix}$$

$$L^T\mathbf{y} = \mathbf{z}$$

$$\begin{bmatrix} 1 & -1 & 2 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \mathbf{y} = \begin{bmatrix} 1 \\ 0 \\ \frac{1}{2} \end{bmatrix}$$

$$\mathbf{y} = \begin{bmatrix} -\frac{1}{2} \\ -\frac{1}{2} \\ \frac{1}{2} \end{bmatrix}$$

University of Washington
Department of Mathematics
MATH 464, Spring 2020

February 23, 2022

Homework 3

Section 2.4

Problem 5:

Let $A = N - P$ where

$$N = \begin{bmatrix} 5 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 5 \end{bmatrix}, P = \begin{bmatrix} -1 & -1 & 1 \\ -2 & -1 & 2 \\ -1 & -1 & -1 \end{bmatrix}$$

(a) Show that the iterative method for this splitting converges to the solution of $A\mathbf{x} = \mathbf{b}$.

(b) Use (2.57) to argue why you expect the Jacobi method to converge faster than this method.

Solution:

(a) We have

$$N^{-1} = \begin{bmatrix} \frac{1}{5} & 0 & 0 \\ 0 & \frac{1}{4} & 0 \\ 0 & 0 & \frac{1}{5} \end{bmatrix}, N^{-1}P = \begin{bmatrix} -\frac{1}{5} & -\frac{1}{5} & \frac{1}{5} \\ -\frac{1}{2} & -\frac{1}{4} & \frac{1}{2} \\ -\frac{1}{5} & -\frac{1}{5} & -\frac{1}{5} \end{bmatrix}$$

The properties of $A = N - P$ tells us that we need to find $\|N^{-1}P\| < 1$ for some matrix norm.

$$\|A\|_\infty = \max_{1 \le i \le n} \sum_{j=1}^{n} |a_{ij}|$$

$$\|N^{-1}P\|_1 = \frac{5}{4}$$

$$\|A\|_1 = \max_{1 \le j \le n} \sum_{i=1}^{n} |a_{ij}|$$

$$\|N^{-1}P\|_1 = \frac{9}{10}$$

Since for the matrix 1 norm $\|N^{-1}P\| < 1$, this splitting converges to the solution of $A\mathbf{x} = \mathbf{b}$.

(b)

$$A = \begin{bmatrix} 6 & 1 & -1 \\ 2 & 5 & -2 \\ 1 & 1 & 6 \end{bmatrix}$$

$$M_J = -\begin{bmatrix} 0 & \frac{1}{6} & -\frac{1}{6} \\ \frac{2}{5} & 0 & -\frac{2}{5} \\ \frac{1}{6} & \frac{1}{6} & 0 \end{bmatrix}$$

Thus we can obtain

$$\|M_J\|_1 = \frac{1}{6} + \frac{2}{5} = \frac{17}{30}$$
$$\|M_J\|_\infty = \frac{2}{5} + \frac{2}{5} = \frac{4}{5}$$

By (2.57), we would expect that if M is 'smaller', which in this case should the matrix norms resemble, the error vector for Jacobi Method should converge faster to zero than this method. We would then expect that the solution vector for the Jacobi Method also converge faster to the true solution than this method.

Problem 7:

How many multiplications and divisions are required for one iteration of the Gauss-Seidel method? How many iterations may be performed before this number exceeds the operations count of Gauss elimination?

Solution:

For one iteration of one entry of solution vector we have the following equation

$$x_i^{(k+1)} = \frac{\left(b_i - \sum_{j<i} a_{ij} x_j^{(k+1)} - \sum_{j>i} a_{ij} x^{(k)}\right)}{a_{ii}}$$

By counting it, we have $n-1+1 = n$ multiplications and divisions. Considering we have n entries of x in total, one iteration of the Gauss-Seidel method requires n^2 multiplications and divisions. We know that Gaussian Elimination has $\frac{1}{3}n^3$ multiplications and divisions, therefore we need the iteration number $k = \frac{n}{3}$ to exceed the operation counts of Gaussian Elimination.

Problem 9:

Let A be an $(m \times n)$ matrix. Using Problem 13, Section 2.1, show that $A^{\mathrm{T}}A$ is an $(n \times n)$ symmetric matrix.

Solution:

From Problem 13, Section 2.1 we have the following:

If A is $(r \times s)$ and B is $(r \times s)$ then $(A+B)^{\mathrm{T}} = A^{\mathrm{T}} + B^{\mathrm{T}}$.
If A is $(r \times s)$ and B is $(s \times p)$ then $(AB)^{\mathrm{T}} = B^{\mathrm{T}}A^{\mathrm{T}}$
$(A^{\mathrm{T}})^{\mathrm{T}} = A$.

Then,

$$\begin{aligned}(A^{\mathrm{T}}A)^{\mathrm{T}} &= A^{\mathrm{T}}(A^{\mathrm{T}})^{\mathrm{T}} \\ &= A^{\mathrm{T}}A\end{aligned}$$

and a matrix B is symmetric if $B^{\mathrm{T}} = B$.

Problem 13:
For the matrix $A^{\mathrm{T}}A$ of Example 2.19, verify that the Jacobi matrix $M_J = -D^{-1}(L+U)$ is given by

$$M_J = -\begin{bmatrix} 0 & \frac{3}{7} \\ \frac{3}{2} & 0 \end{bmatrix} \text{ and } M_J^2 = \begin{bmatrix} \frac{9}{14} & 0 \\ 0 & \frac{9}{14} \end{bmatrix}.$$

Using $M_J^4 = M_J^2 M_J^2, M_J^6 = M_J^2 M_J^2 M_J^2$, etc., show that the entries of M_J^{2n} all tend to zero as $n \to \infty$. From this information, give a bound for $\|M_J^i\|_\infty$; and using (2.57), show that the Jacobi method must converge for any initial guess $\mathbf{x}^{(0)}$. Note that $A^{\mathrm{T}}A$ is not diagonally dominant so that Theorem 2.3 does not apply. Additionally, $\|M_J\| > 1$ for the three matrix norms of (2.41) so that Theorem 2.2 does not apply either. However, the analysis of this problem shows why the Jacobi method is convergent in Problem 3.

Solution:
From Example 2.19 we have

$$A^{\mathrm{T}}A = \begin{bmatrix} 14 & 6 \\ 6 & 4 \end{bmatrix}$$

Then,

$$\begin{aligned} M_J &= \begin{bmatrix} 0 & \frac{6}{14} \\ \frac{6}{4} & 0 \end{bmatrix} \\ &= \begin{bmatrix} 0 & \frac{3}{7} \\ \frac{3}{2} & 0 \end{bmatrix} \\ M_J^2 &= \begin{bmatrix} \frac{9}{14} & 0 \\ 0 & \frac{9}{14} \end{bmatrix} \end{aligned}$$

We can easily see that even exponents of M_J we have

$$M_J^{2n} = \begin{bmatrix} \left(\frac{9}{14}\right)^n & 0 \\ 0 & \left(\frac{9}{14}\right)^n \end{bmatrix}$$

where

$$\|M_J^{2n}\|_\infty = \left(\frac{9}{14}\right)^n.$$

Moreover, by the definition of matrix norm we know that for odd exponents of M_J we have

$$\begin{aligned}\|M_J^{2n+1}\|_\infty &\leq \|M_J^{2n}\|_\infty \|M_J\|_\infty \\ &\leq \frac{3}{2}\left(\frac{9}{14}\right)^n\end{aligned}$$

Since $\lim_{k\to\infty} \|M_J^k\|_\infty = 0$ in all cases, the error vector

$$\|\mathbf{e}^{(k+1)}\|_\infty \leq \|M_J^k\|_\infty \|\mathbf{e}_\infty^{(0)}$$

would also approach 0 as $k \to \infty$, hence the Jacobi method is convergent in this problem.

Section 2.5

Problem 1:
Use the methods of this section to find the best least-squares solution of the system

$$\begin{aligned} x_1 + x_2 &= 1 \\ 2x_1 + x_2 &= 0 \\ x_1 - x_2 &= 0 \end{aligned}$$

Solution:

$$\begin{aligned} A &= \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 1 & -1 \end{bmatrix} \\ A^{\mathrm{T}} &= \begin{bmatrix} 1 & 2 & 1 \\ 1 & 1 & -1 \end{bmatrix} \\ A^{\mathrm{T}}A &= \begin{bmatrix} 6 & 2 \\ 2 & 3 \end{bmatrix} \\ A^{\mathrm{T}}\mathbf{b} &= \begin{bmatrix} 1 \\ 1 \end{bmatrix} \\ \begin{bmatrix} 6 & 2 \\ 2 & 3 \end{bmatrix} \mathbf{x}^* &= \begin{bmatrix} 1 \\ 1 \end{bmatrix} \\ \mathbf{x}^* &= \begin{bmatrix} \frac{1}{14} \\ \frac{2}{7} \end{bmatrix} \end{aligned}$$

Problem 3:
Consider the (3×3) system

$$\begin{aligned} x_1 + 3x_2 + 7x_3 &= 1 \\ 2x_1 + x_2 - x_3 &= 1 \\ x_1 + 2x_2 + 4x_3 &= 1 \end{aligned}$$

Show the system has no solution (so that coefficient matrix A must be singular). Next find the best least-squares solution by the techniques of this section. Note that even though $A^{\mathrm{T}}A$ is singular, the equation corresponding to (2.73) is solvable.

Solution:

We can find the nullspace of A.

$$\begin{bmatrix} 1 & 3 & 7 & 0 \\ 2 & 1 & -1 & 0 \\ 1 & 2 & 4 & 0 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 3 & 7 & 0 \\ 0 & -5 & -15 & 0 \\ 1 & 2 & 4 & 0 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 3 & 7 & 0 \\ 0 & -5 & -15 & 0 \\ 0 & -1 & -3 & 0 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 3 & 7 & 0 \\ 0 & 1 & 3 & 0 \\ 0 & 1 & 3 & 0 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & -2 & 0 \\ 0 & 1 & 3 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Therefore A is singular and $A\mathbf{x} = \mathbf{b}$ has no solution.

$$A^{\mathrm{T}} = \begin{bmatrix} 1 & 2 & 1 \\ 3 & 1 & 2 \\ 7 & -1 & 4 \end{bmatrix}$$

$$A^{\mathrm{T}}A = \begin{bmatrix} 6 & 7 & 9 \\ 7 & 14 & 28 \\ 9 & 28 & 66 \end{bmatrix}$$

$$A^{\mathrm{T}}\mathbf{b} = \begin{bmatrix} 4 \\ 6 \\ 10 \end{bmatrix}$$

To solve $A^{\mathrm{T}}A\mathbf{x} = A^{\mathrm{T}}\mathbf{b}$, we do gaussian elimination

$$\begin{bmatrix} 6 & 7 & 9 & 4 \\ 7 & 14 & 28 & 6 \\ 9 & 28 & 66 & 10 \end{bmatrix}$$

$$\begin{bmatrix} 9 & 28 & 66 & 10 \\ 7 & 14 & 28 & 6 \\ 6 & 7 & 9 & 4 \end{bmatrix}$$

$$\begin{bmatrix} 9 & 28 & 66 & 10 \\ 0 & -\frac{70}{9} & -\frac{70}{3} & -\frac{16}{9} \\ 6 & 7 & 9 & 4 \end{bmatrix}$$

$$\begin{bmatrix} 9 & 28 & 66 & 10 \\ 0 & -\frac{70}{9} & -\frac{70}{3} & -\frac{16}{9} \\ 0 & -\frac{35}{3} & -35 & -\frac{8}{3} \end{bmatrix}$$

$$\begin{bmatrix} 9 & 28 & 66 & 10 \\ 0 & -\frac{35}{3} & -35 & -\frac{8}{3} \\ 0 & -\frac{70}{9} & -\frac{70}{3} & -\frac{16}{9} \end{bmatrix}$$

$$\begin{bmatrix} 9 & 28 & 66 & 10 \\ 0 & -\frac{35}{3} & -35 & -\frac{8}{3} \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Let $x_3 = a$.

$$\begin{aligned} 9x_1 + 28x_2 + 66a &= 10 \\ 35x_2 + 105a &= 8 \end{aligned}$$

$$\begin{aligned} 9x_1 + 28x_2 + 66a &= 10 \\ x_2 &= \frac{8}{35} - 3a \end{aligned}$$

$$\begin{aligned} 9x_1 + 28\left(\frac{8}{35} - 3a\right) + 66a &= 10 \\ x_2 &= \frac{8}{35} - 3a \end{aligned}$$

$$\begin{aligned} 9x_1 + \frac{224}{35} - 84a + 66a &= 10 \\ x_2 &= \frac{8}{35} - 3a \end{aligned}$$

$$\begin{aligned} 9x_1 - 18a &= 10 - \frac{224}{35} \\ x_2 &= \frac{8}{35} - 3a \end{aligned}$$

$$\begin{aligned} 9x_1 &= \frac{126}{35} + 18a \\ x_2 &= \frac{8}{35} - 3a \end{aligned}$$

$$\begin{aligned} x_1 &= \frac{14}{35} + 2a \\ x_2 &= \frac{8}{35} - 3a \end{aligned}$$

Thus

$$\mathbf{x}^* = \begin{bmatrix} \frac{14}{35} + 2a \\ \frac{8}{35} - 3a \\ a \end{bmatrix}$$

Problem 4:

For the matrix A in Problem 3, find a nonzero vector $\mathbf{x}$ such that $\mathbf{x}^{\mathrm{T}}(A^{\mathrm{T}}A)\mathbf{x} = 0$, and thus conclude that $A^{\mathrm{T}}A$ cannot be positive-definite.

Solution:

We have already found the nullspace of A. Notice that

$$\begin{aligned} \mathbf{x}^{\mathrm{T}}(A^{\mathrm{T}}A)\mathbf{x} &= 0 \\ (A\mathbf{x})^{\mathrm{T}}(A\mathbf{x}) &= 0 \end{aligned}$$

The nullspace of A is

$$\begin{bmatrix} 1 & 0 & -2 & 0 \\ 0 & 1 & 3 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Let $x_3 = a$.

$$\mathbf{x} = \begin{bmatrix} 2a \\ -3a \\ 3a \end{bmatrix}$$

For any $a \neq 0$ we have a nonzero vector $\mathbf{x}$ that makes the above equation equal to 0. Thus $A^{\mathrm{T}}A$ cannot be positive-definite.

Problem 5a:

If $y = a_0x + a_1$, choose a_0 and a_1 such that this straight line is the best least-squares fit to these (x, y) data points:

(a) (1,1),(4,2),(8,4),11,5)

(b) (-1,0),(0,1),(1,2),(2,4)

(c) (-2,2),(-1,1),(1,0),(2,-1)

Solution:

(a)

$$A = \begin{bmatrix} 1 & 1 \\ 4 & 1 \\ 8 & 1 \\ 11 & 1 \end{bmatrix}$$

$$\mathbf{b} = \begin{bmatrix} 1 \\ 2 \\ 4 \\ 5 \end{bmatrix}$$

$$A^{\mathrm{T}}A = \begin{bmatrix} 202 & 24 \\ 24 & 4 \end{bmatrix}$$

$$A^{\mathrm{T}}\mathbf{b} = \begin{bmatrix} 96 \\ 12 \end{bmatrix}$$

Use Gaussian Elimination

$$\begin{bmatrix} 202 & 24 & 96 \\ 24 & 4 & 12 \end{bmatrix}$$
$$\begin{bmatrix} 202 & 24 & 96 \\ 0 & \frac{116}{101} & \frac{60}{101} \end{bmatrix}$$
$$\begin{bmatrix} 202 & 24 & 96 \\ 0 & 1 & \frac{15}{29} \end{bmatrix}$$
$$\begin{bmatrix} 202 & 0 & \frac{2424}{29} \\ 0 & 1 & \frac{15}{29} \end{bmatrix}$$
$$\begin{bmatrix} 1 & 0 & \frac{12}{29} \\ 0 & 1 & \frac{15}{29} \end{bmatrix}$$

Hence,

$$y = \frac{12}{29}x + \frac{15}{29}$$

(b)

$$A = \begin{bmatrix} -1 & 1 \\ 0 & 1 \\ 1 & 1 \\ 2 & 1 \end{bmatrix}$$
$$\mathbf{b} = \begin{bmatrix} 0 \\ 1 \\ 2 \\ 4 \end{bmatrix}$$
$$A^{\mathrm{T}}A = \begin{bmatrix} 6 & 2 \\ 2 & 4 \end{bmatrix}$$
$$A^{\mathrm{T}}\mathbf{b} = \begin{bmatrix} 10 \\ 7 \end{bmatrix}$$

Use Gaussian Elimination

$$\begin{bmatrix} 6 & 2 & 10 \\ 2 & 4 & 7 \end{bmatrix}$$
$$\begin{bmatrix} 6 & 2 & 10 \\ 0 & \frac{10}{3} & \frac{11}{3} \end{bmatrix}$$
$$\begin{bmatrix} 6 & 2 & 10 \\ 0 & 1 & \frac{11}{10} \end{bmatrix}$$
$$\begin{bmatrix} 6 & 0 & \frac{39}{5} \\ 0 & 1 & \frac{11}{10} \end{bmatrix}$$
$$\begin{bmatrix} 1 & 0 & \frac{13}{10} \\ 0 & 1 & \frac{11}{10} \end{bmatrix}$$

Hence,

$$y = \frac{13}{10}x + \frac{11}{10}$$

(c)

$$A = \begin{bmatrix} -2 & 1 \\ -1 & 1 \\ 1 & 1 \\ 2 & 1 \end{bmatrix}$$

$$\mathbf{b} = \begin{bmatrix} 2 \\ 1 \\ 0 \\ -1 \end{bmatrix}$$

$$A^{\mathrm{T}}A = \begin{bmatrix} 10 & 0 \\ 0 & 4 \end{bmatrix}$$

$$A^{\mathrm{T}}\mathbf{b} = \begin{bmatrix} -7 \\ 2 \end{bmatrix}$$

Use Gaussian Elimination

$$\begin{bmatrix} 10 & 0 & -7 \\ 0 & 4 & 2 \end{bmatrix}$$

$$\begin{bmatrix} 10 & 0 & -7 \\ 0 & 1 & \frac{1}{2} \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & -\frac{7}{10} \\ 0 & 1 & \frac{1}{2} \end{bmatrix}$$

Hence,

$$y = -\frac{7}{10}x + \frac{1}{2}$$

Problem 6a:

If $y = a_0x^2 + a_1x + a_2$, choose a_0, a_1, and a_2 such that this quadratic is the best least-squares fit to these (x, y) data points:

(a) (-2,2),(-1,1),(1,1),(2,2)

Solution:

$$A = \begin{bmatrix} 4 & -2 & 1 \\ 1 & -1 & 1 \\ 1 & 1 & 1 \\ 4 & 2 & 1 \end{bmatrix}$$

$$\mathbf{b} = \begin{bmatrix} 2 \\ 1 \\ 1 \\ 2 \end{bmatrix}$$

$$A^{\mathrm{T}}A = \begin{bmatrix} 34 & 0 & 10 \\ 0 & 10 & 0 \\ 10 & 0 & 4 \end{bmatrix}$$

$$A^{\mathrm{T}}\mathbf{b} = \begin{bmatrix} 18 \\ 0 \\ 6 \end{bmatrix}$$

Use Gaussian Elimination

$$\begin{bmatrix} 34 & 0 & 10 & 18 \\ 0 & 10 & 0 & 0 \\ 10 & 0 & 4 & 6 \end{bmatrix}$$

$$\begin{bmatrix} 34 & 0 & 10 & 18 \\ 0 & 10 & 0 & 0 \\ 0 & 0 & \frac{18}{17} & \frac{12}{17} \end{bmatrix}$$

$$\begin{bmatrix} 34 & 0 & 10 & 18 \\ 0 & 10 & 0 & 0 \\ 0 & 0 & 1 & \frac{2}{3} \end{bmatrix}$$

$$\begin{bmatrix} 34 & 0 & 0 & \frac{34}{3} \\ 0 & 10 & 0 & 0 \\ 0 & 0 & 1 & \frac{2}{3} \end{bmatrix}$$

$$\begin{bmatrix} 34 & 0 & 0 & \frac{34}{3} \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & \frac{2}{3} \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & 0 & \frac{1}{3} \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & \frac{2}{3} \end{bmatrix}$$

Hence,

$$y = \frac{1}{3}x^2 + \frac{2}{3}$$

University of Washington
Department of Mathematics
MATH 464, Spring 2020

February 23, 2022

Homework 4

Section 4.3.1

Problem 1:
The equation $x^3 - 13x + 18 = 0$ is equivalent to the fixed-point problem $x = g(x)$ for each of these choices:

(a) $g(x) = (x^3 + 18)/13$

(b) $g(x) = (13x - 18)^{\frac{1}{3}}$

(c) $g(x) = (13x - 18)/x^2$

(d) $g(x) = x^3 - 12x + 18$

Show by direct substitution that $s = 2$ is a fixed point in each case above. For which choices is $|g'(x)| < 1$?

Solution:

(a)

$$g(2) = \frac{(2^3 + 18)}{13}$$
$$g(2) = \frac{26}{13}$$
$$g(2) = 2$$

Then we find the derivative

$$g'(x) = \frac{3}{13}x^2$$
$$|g'(2)| = \frac{12}{13} < 1$$

(b)

$$g(2) = (13 \cdot 2 - 18)^{\frac{1}{3}}$$
$$g(2) = 8^{\frac{1}{3}}$$
$$g(2) = 2$$

Then we find the derivative

$$g'(x) = \frac{13}{3(13x-18)^{\frac{2}{3}}}$$
$$|g'(2)| = \frac{13}{12} > 1$$

(c)

$$g(2) = \frac{(13 \cdot 2 - 18)}{2^2}$$
$$g(2) = \frac{8}{4}$$
$$g(2) = 2$$

Then we find the derivative

$$g'(x) = -\frac{13}{x^2} + \frac{36}{x^3}$$
$$|g'(s)| = \frac{5}{4} > 1$$

(d)

$$g(2) = 2^3 - 12 \cdot 2 + 18$$
$$g(2) = 8 - 24 + 18$$
$$g(2) = 2$$

Then we find the derivative

$$g'(x) = 3x^2 - 12$$
$$|g'(2)| = 0 < 1$$

From the graph comparing different $g'(x)$ (or checking the second derivative) we can see that $g'(x)$ from (a) satisfies

$$|g'(2)| < 1, x \in (-2.082, 2.082)$$

while $g'(x)$ from (d) only satisfies

$$|g'(2)| < 1, x \in (1.915, 2.082)$$

Clearly $g'(x)$ from (a) has a larger interval of $|g'(2)| < 1$ that contains $x = 2$.

Problem 4:

Verify that the equation $x^2 - c = 0 (c > 0)$ is equivalent to the fixed-point problem $x = (x^2 + c)/2x$. One fixed point is $s = \sqrt{c}$; verify that $0 < g'(x) < 1$ for $\sqrt{c} < x < \infty$. By Problem 5 below, the fixed-point iteration will converge for any x_0 in $(\sqrt{c}, \infty)$. Set $x_0 = c$ and execute six steps of the iteration for $c = 3, 5, 7$. Compare your estimates with the actual solution.

Solution:

To validate, we check the fixed-point iteration function

$$\begin{aligned} g(x) &= \frac{x^2+c}{2x} \\ x &= \frac{x^2+c}{2x} \\ 2x^2 &= x^2+c \\ x^2-c &= 0 \end{aligned}$$

To verify the interval bound of $0 < g'(x) < 1$, we first find the derivative of $g(x)$

$$\begin{aligned} g(x) &= \frac{x^2+c}{2x} \\ &= \frac{x^2}{2x} + \frac{c}{2x} \\ g'(x) &= \frac{1}{2} - \frac{c}{2x^2} \end{aligned}$$

Then substitute to find the interval

$$\begin{aligned} 0 < g'(x) &< 1 \\ 0 < \frac{1}{2} - \frac{c}{2x^2} &< 1 \\ 0 < 1 - \frac{c}{x^2} &< 2 \\ -2 < \frac{c}{x^2} - 1 &< 0 \\ -1 < \frac{c}{x^2} &< 1 \\ \left|\frac{c}{x^2}\right| &< 1 \\ \frac{c}{x^2} &< 1 \\ \frac{x^2}{c} &> 1 \\ \sqrt{c} < x &< \infty \end{aligned}$$

Check the following table

c	3	5	7
x_0	3	5	7
x_1	$g(x_0) = 2$	$g(x_0) = 3$	$g(x_0) = 4$
x_2	$g(x_1) = 1.75$	$g(x_1) = 2.33333333333$	$g(x_1) = 2.875$
x_3	$g(x_2) = 1.73214285714$	$g(x_2) = 2.2380952381$	$g(x_2) = 2.65489130435$
x_4	$g(x_3) = 1.73205081001$	$g(x_3) = 2.23606889564$	$g(x_3) = 2.64576704419$
x_5	$g(x_4) = 1.73205080757$	$g(x_4) = 2.2360679775$	$g(x_4) = 2.64575131111$
x_6	$g(x_5) = 1.73205080757$	$g(x_5) = 2.2360679775$	$g(x_5) = 2.64575131106$
$\sqrt{c}$	1.73205080757	2.2360679775	2.64575131106

We see that $x_0 = c$ with six steps of the iteration for $c = 3, 5, 7$ on the fix-point iteration $g(x) = \frac{x^2+c}{2x}$, the estimates converges to the actual solution $\sqrt{c}$ quickly and matches up to 12 significant digits.

Problem 6:

Evaluate: $s = \sqrt[3]{6 + \sqrt[3]{6 + \sqrt[3]{6 + \cdots}}}$. [Hint: Let $x_0 = 0$ and consider $g(x) = \sqrt[3]{6 + x}$].

Solution:

$s = \sqrt[3]{6 + \sqrt[3]{6 + \sqrt[3]{6 + \cdots}}}$ is the limit of the sequence $x_{n+1} = g(x_n), g(x) = \sqrt[3]{6 + x}$ as $n \to \infty$.
We easily find $s = 2$ as a fix-point of $g(x)$

$$2 = \sqrt[3]{8} = \sqrt[3]{6 + 2}$$
$$s = g(s)$$

We also find the first and second derivative of $g(x)$

$$g(x) = \sqrt[3]{6 + x}$$
$$g'(x) = \frac{1}{3(6 + x)^{\frac{2}{3}}}$$
$$g''(x) = -\frac{2}{9\,(6 + x)^{\frac{5}{3}}}$$

We want to use Theorem 4.2 and 4.3, hence we want to satisfy that $g(I) \subseteq I$, and $|g'(x)| \leq L < 1$.
For the interval $[0, 2]$, we find $g(0) = \sqrt[3]{6}, g(2) = 2, g'(0) = \frac{1}{3\sqrt[3]{36}}, g'(2) = \frac{1}{12}$.
We also find that $g''(x) < 0, x \in [0, 2]$.
This implies that the $g'(x)$ is decreasing on $[0, 2]$, and $0 < g'(x) \leq g'(0) < 1$.
Since $g'(x) > 0$ on interval $[0, 2]$, we can also conclude that $g(I) \subseteq I, I = [0, 2]$.
Thereby Theorem 4.3 we have verified that $g(x) = \sqrt[3]{6 + x}$ does converge to the a fix-point over the interval $[0, 2]$ with $x_0 = 0$, and since we have already found one fix-point $s = 2$, by Theorem 4.2, fix-point iteration on $g(x)$ will converge to the unique fix-point $s = 2$ with $x_0 = 0$.

Problem 9:

Let $g(x) = x^2$. From a graph(as in Problem 8), deduce for what values $x_0, -\infty < x_0 < \infty$, the iteration $x_{i+1} = g(x_i)$ will converge to a fixed point of $g(x)$ and for what values x_0 will the iteration diverge.

Solution:

We can first find a conservative bound by trying to find an interval with

$$|g'(x)| \leq L < 1$$

for some L and apply Theorem 4.3. The derivative of $g(x)$ is

$$g'(x) = 2x$$

Clearly for values $-\frac{1}{2} < x < \frac{1}{2}, |g'(x)| \leq L < 1$, the fix-point iteration converges to 0.
However, for this iteration $x_{i+1} = g(x_i)$, it is equivalent to find

$$x_n = \lim_{n \to \infty} x_0^{2^n}$$

for different x_0. It is obvious that for $x_0 = -1$ or $x_0 = 1$, the sequence stays at 1.
For $-1 < x_0 < 1$, the sequence converges to 0.
Thus, we should be reminded that Theorem 4.3 is only sufficient to find a convergence over a strict interval for a given fix-point iteration. It is not necessary for convergence.

Section 4.3.3

Problem 2:

In Problem 8 it is shown that the errors in Newton's method satisfy $e_{n+1} \approx Ke_n^2$ where

$$K = f''(s)/2f'(s)$$

[under the assumption that $f(s) = 0$ and $f'(s) \neq 0$]. For $f(x)$ in Problem 1, verify that $K = 1$.

Solution:

In Problem 1 we have

$$f(x) = x^3 - 2x^2 + 2x - 1$$

Given the root of $f(x)$ is $s = 1$,

$$\begin{aligned} f'(x) &= 3x^2 - 4x + 2 \\ f''(x) &= 6x - 4 \\ K &= \frac{f''(s)}{2f'(s)} \\ &= \frac{f''(1)}{2f'(1)} \\ &= \frac{6 \cdot 1 - 4}{2(3 \cdot 1^2 - 4 \cdot 1 + 2)} \\ &= \frac{2}{2} \\ &= 1 \end{aligned}$$

Problem 7:

Prove that the tangent line to the graph of $y = f(x)$ at the point $(x_n, f(x_n))$ intersects the x-axis when $x = x_n - f(x_n)/f'(x_n)$.

Solution:

The function of the tangent line at the point $(x_n, f(x_n))$ is

$$\begin{aligned} y - f(x_n) &= f'(x_n)(x - x_n) \\ y &= f(x_n) + f'(x_n)(x - x_n) \end{aligned}$$

Let $x = x_n - \frac{f(x_n)}{f'(x_n)}$,

$$\begin{aligned} y &= f(x_n) + f'(x_n)\left(x_n - \frac{f(x_n)}{f'(x_n)} - x_n\right) \\ &= f(x_n) - f'(x_n)\frac{f(x_n)}{f'(x_n)} \\ &= f(x_n) - f(x_n) \\ &= 0 \end{aligned}$$

This confirms that the tangent line y intersects the x-axis, which means $y = 0$ at point $(x_n, f(x_n))$.

Problem 12:

As an extreme case of a function for which Newton's method is slowly convergent, consider $f(x) = (x-\alpha)^n$ for n some positive integer and α some real number. Show that Newton's method generates the sequence

$$x_{i+1} = (1 - 1/n)x_i + \alpha/n,$$

and then show that $x_{i+1} - \alpha = (1 - 1/n)(x_i - \alpha)$. This is a special case of Problem 9 above.

Solution:

Using Newton's method, we have

$$\begin{aligned} x_{i+1} &= x_i - \frac{(x_i - \alpha)^n}{n(x_i - \alpha)^{n-1}} \\ &= x_i - \frac{x_i - \alpha}{n} \\ &= x_i - \frac{1}{n}x_i + \frac{a}{n} \\ &= \left(1 - \frac{1}{n}\right)x_i + \frac{a}{n} \end{aligned}$$

We can then solve for $x_{i+1} - \alpha$

$$\begin{aligned} x_{i+1} - \alpha &= \left(1 - \frac{1}{n}\right)x_i + \frac{a}{n} - \alpha \\ &= \left(1 - \frac{1}{n}\right)x_i + \frac{1}{n}\alpha - \alpha \\ &= \left(1 - \frac{1}{n}\right)x_i - \left(1 - \frac{1}{n}\right)\alpha \\ &= \left(1 - \frac{1}{n}\right)(x_i - \alpha) \end{aligned}$$

University of Washington
Department of Mathematics
MATH 464, Spring 2020

February 23, 2022

Homework 5

Section 4.4.1

Problem 2:

Let $p(x)$ be given by (4.12) and assume that none of its coefficients is zero. Show that evaluation of $p(\alpha)$ by direct substitution requires at least $2n-1$ multiplications.

Solution:

Eq (4.12) is a nth degree polynomial

$$p(x) = a_0x^n + a_1x^{n-1} + \cdots + a_{n-1}x + a_n, a_0 \neq 0$$

Assuming none of its coefficients is zero, we evaluate of $p(\alpha)$ by direct substitution.
Define $m(n)$ to be the number of multiplication for the nth power term of α.
We start from $a_{n-1}\alpha$, since a_n has no multiplication.
We see that

$$\begin{aligned}
m(1) &= 1, a_{n-1} \cdot \alpha. && \leftarrow \text{ save } \alpha \\
m(2) &= 2, a_{n-2} \cdot \alpha \cdot \alpha. && \leftarrow \text{ save } \alpha^2, \text{ use } \alpha \\
m(3) &= 2, a_{n-3} \cdot \alpha \cdot \alpha^2. && \leftarrow \text{ save } \alpha^3, \text{ use } \alpha^2 \\
m(4) &= 2, a_{n-4} \cdot \alpha \cdot \alpha^3 && \leftarrow \text{ save } \alpha^4, \text{ use } \alpha^3 \\
&\vdots \\
m(n-1) &= 2, a_1 \cdot \alpha \cdot \alpha^{n-1} && \leftarrow \text{ save } \alpha^{n-1}, \text{ use } \alpha^{n-2} \\
m(n) &= 2, a_0 \cdot \alpha \cdot \alpha^{n-1} && \leftarrow \text{ save } \alpha^n, \text{ use } \alpha^{n-1}
\end{aligned}$$

Thus the evaluation of $p(\alpha)$ by direction substitution requires at least

$$\begin{aligned}
\sum_{i=1}^{n} m(i) &= m(1) + \sum_{i=2}^{n} m(i) \\
&= 1 + \sum_{i=2}^{n} 2 \\
&= 1 + (n-1) \cdot 2 \\
&= 2n - 2 + 1 \\
&= 2n - 1
\end{aligned}$$

multiplications.

Problem 3:

Establish that the number, b_n, generated by the synthetic division algorithm satisfies $b_n = p(\alpha)$. [Hint: In (4.16), let $p(x) = P(x)$ and $Q(x) = (x-\alpha)$. Let the coefficients of $Q(x)$ be $b_0, b_1, \cdots, b_{n-1}$ and equate like powers on both sides of (4.16).]

Solution:

To show that the b_n generated by the synthetic division algorithm satisfies $b_n = p(\alpha)$ we need to show that b_n is unique and in different forms of $p(x)$ equates to the same.
Define $b_0 = a_0$ and $b_j = \alpha b_{j-1} + a_j, 1 \leq j \leq n$ for the nested multiplication method

$$p(x) = x(x(x(a_0x + a_1) + a_2) + a_3) + \cdots) + a_n$$

It is easy to see that since b_j are the temporary partial result of the doing nested multiplication algorithm on $p(\alpha)$,

$$p(\alpha) = b_n$$

Then, define c_j to be the coefficients of $q_{n-1}(x)$

$$q_{n-1} = c_0x^{n-1} + c_1x^{n-2} + \cdots + c_{n-1}$$

in

$$p(x) = (x - \alpha)q_{n-1}(x) + r_0(x)$$

and $c_n = r_0(x)$. By substitution, the synthetic division form of $p(x)$ expands into

$$\begin{aligned} p(x) &= (x - \alpha)(c_0x^{n-1} + c_1x^{n-2} + \cdots + c_{n-1}) + c_n \\ &= c_0x^n + c_1x^{n-1} + \cdots + c_{n-1}x - \alpha c_0x^{n-1} - \alpha c_1x^{n-2} - \cdots - \alpha c_{n-1} + c_n \\ &= c_0x^n + (c_1 - \alpha c_0)x^{n-1} + \cdots + (c_{n-1} - \alpha c_{n-2})x + c_n \end{aligned}$$

We know that the default form of $p(x)$ is

$$p(x) = a_0x^n + a_1x^{n-1} + \cdots + a_{n-1}x + a_n, a \neq 0$$

Thus we can clearly see that $c_0 = a_0$

$$c_j = \alpha c_{j-1} + a_j, 1 \leq j \leq n$$

which has the same form as the b_j we defined. Notice that α, a_0 are determined, which uniquely determines c_j and b_j.
Thus it is clear by induction that $c_j = b_j$, and thus, the c_n coefficient of $q_{n-1}(x)$ in $p(x)$ in synthetic division form also evaluate to $p(\alpha) = c_n$.

Problem 4:

Start with Eq. (4.23) and show that $r_2 = \frac{p''(a)}{2}$ where r_2 is obtained in the same manner as r_0 and r_1.

Solution:
(4.23)

$$p(x) = (x-\alpha)^2 q_{n-2}(x) + r_1(x-\alpha) + r_0$$

is obtained by substituting

$$q_{n-1}(x) = (x-\alpha)q_{n-2}(x) + r_1, r_1 = q_{n-1}(\alpha)$$

into

$$p(x) = (x-\alpha)q_{n-1} + r_0(x)$$

We could apply the same procedure to obtain r_2

$$q_{n-2}(x) = (x-\alpha)q_{n-3}(x) + r_2, r_2 = q_{n-2}(\alpha)$$

and thus (4.21) becomes

$$p(x) = (x-\alpha)^3 q_{n-3}(x) + r_2(x-\alpha)^2 + r_1(x-\alpha) + r_0 \tag{1}$$

Differentiating (1) twice we have

$$p''(\alpha) = 2r_2$$
$$r_2 = \frac{p''(a)}{2}$$

Thus we have shown $r_2 = \frac{p''(a)}{2}$.

Section 5.1

Problem 4:
The $sine - integral$,Si(x), occurs frequently in certain applied problems where $x > 0$ and Si(x) is defined by

$$\text{Si}(x) = \int_0^x \frac{\sin(t)}{t} dt$$

One way to estimate Si(x) is to take the truncated kth-degree Taylor's series expansion $p_k(t)$, for $f(t) = \sin(t)$ with $a = 0$, and use

$$\int_0^x \frac{p_k(t)}{t} dt$$

as an approximation to Si(x). How large must k be in order that

$$\left| \int_0^4 \frac{p_k(t)}{t} dt - \text{Si}(4) \right| \leq 10^{-6} \tag{2}$$

To bound the error, use the fact that if $|h(t)| \leq |q(t)|$ for $0 \leq t \leq x$, then $\int_0^x |h(t)|dt \leq \int_0^x |q(t)|dt$.

Solution:
We know that for $x \in [a, b]$, the point-wise upper bound given by Taylor polynomials is

$$|f(x) - p_k(x)| \leq \frac{M_{k+1}}{(k+1)!}|(x-c)^{k+1}|$$

where $M_{k+1} = \max_{a \leq x \leq b} |f^{(k+1)}(x)|$.
We want

$$\left|\int_0^4 \frac{p_k(t)}{t}dt - \mathrm{Si}(4)\right| \leq 10^{-6}$$
$$\left|\int_0^4 \frac{p_k(t)}{t}dt - \int_0^4 \frac{\sin(t)}{t}dt\right| \leq 10^{-6}$$
$$\left|\int_0^4 \frac{p_k(t) - \sin(t)}{t}dt\right| \leq 10^{-6}$$

Notice that $\sin(x) \leq 1$ for all x. Thus $M_{n+1} = 1$, and the point-wise upper bound evaluate to

$$|\sin(x) - p_k(x)| \leq \frac{1}{(k+1)!}|t^{k+1}|$$

for $p_k(t)$ of $\sin(t)$ centered at 0. Thus

$$\left|\int_0^4 \frac{\frac{1}{(k+1)!}t^{k+1}}{t}dt\right| \leq 10^{-6}$$
$$\left|\int_0^4 \frac{t^k}{(k+1)!}dt\right| \leq 10^{-6}$$

Evaluating the integral we have

$$\left|\frac{4^{k+1}}{(k+1)!\,(k+1)}\right| \leq 10^{-6} \tag{3}$$

Plotting the function we can easily check that for $k = 17$, (3) evaluates to 5.9630207217×10^7 which is clearly less than 10^-6, while $k = 16$, (3) evaluates to $2.84120399094 \times 10^{-6}$. Hence we can conclude that k must be at least 17 in order to make the inequality (2) be true.

Section 5.2.1

Problem 1a:
The Lagrange Interpolating Polynomial is given by

$$p(x) = \sum_{j=0}^{n} y_j l_j(x)$$

where

$$l_j = \prod_{\substack{i=0 \\ i \neq j}}^{n} \frac{(x - x_i)}{(x_j - x_i)}$$

Thus we find each l_j

$$\begin{aligned}
l_0 x &= \frac{(x-0)(x-2)(x-3)}{(-1-0)(-1-2)(-1-3)} = \frac{x^3 - 5x^2 + 6x}{-12} \\
l_1 x &= \frac{(x-(-1))(x-2)(x-3)}{(0-(-1))(0-2)(0-3)} = \frac{x^3 - 4x^2 + x + 6}{6} \\
l_2 x &= \frac{(x-(-1))(x-0)(x-3)}{(2-(-1))(2-0)(2-3)} = \frac{x^3 - 2x^2 - 3x}{-6} \\
l_3 x &= \frac{(x-(-1))(x-0)(x-2)}{(3-(-1))(3-0)(3-2)} = \frac{x^3 - x^2 - 2x}{12}
\end{aligned}$$

Then find $p(x)$

$$\begin{aligned}
p(x) &= \sum_{j=0}^{n} y_j l_j(x) \\
&= y_0 \frac{x^3 - 5x^2 + 6x}{-12} + y_1 \frac{x^3 - 4x^2 + x + 6}{6} + y_2 \frac{x^3 - 2x^2 - 3x}{-6} + y_3 \frac{x^3 - x^2 - 2x}{12} \\
&= -1 \frac{x^3 - 5x^2 + 6x}{-12} + 3 \frac{x^3 - 4x^2 + x + 6}{6} + 11 \frac{x^3 - 2x^2 - 3x}{-6} + 27 \frac{x^3 - x^2 - 2x}{12}
\end{aligned}$$

Thus

$$l_j(-2) = -13$$

Problem 3a:

The table is constructed below. Thus

$$\begin{aligned} p(x) =& f(x_0) + f[x_0, x_1](x - x0) + f[x_0, x_1, x_2](x - x_0)(x - x_1) + \\ & f[x_0, x_1, x_2, x_3](x - x_0)(x - x_1)(x - x_2) \\ p(-2) =& -1 + 4(-2) + 1 + 0 + (-2 + 1)(-2 - 0)(-2 - 2) \\ =& -13 \end{aligned}$$

Table 1: Divided Difference table

-1	-1			
		4		
0	3		0	
		4		1
2	11		4	
		16		
3	27			

Problem 4a:

The modified table is shown below Thus

Table 2: Divided Difference table

-1	-1				
		4			
0	3		0		
		4		1	
2	11		4		-1.225
		16		-5.125	
3	27		-16.5		
		-17			
4	10				

$$\begin{aligned} q(x) =& f(x_0) + f[x_0, x_1](x - x0) + f[x_0, x_1, x_2](x - x_0)(x - x_1) + \\ & f[x_0, x_1, x_2, x_3](x - x_0)(x - x_1)(x - x_2) + \\ & f[x_0, x_1, x_2, x_3, x_4](x - x_0)(x - x_1)(x - x_2)(x - x_3) \\ q(-2) =& -1 + 4(-2) + 1 + 0 + (-2 + 1)(-2 - 0)(-2 - 2) + \\ & (-1.225)(-2 + 1)(-2 - 0)(-2 - 2)(-2 - 3) \\ =& -62 \end{aligned}$$

Problem 11:

Suppose that P and Q are both sets of $k + 1$ points where $P \cap Q$ has k points. Suppose that $p(x)$ in $\mathcal{P}_k$ interpolates $f(x)$ at the points of P, and $q(x)$ in $\mathcal{P}_k$ interpolates $f(x)$ at the points of Q. Let α denote the only point in $P - (P \cap Q)$; and β, the only point in $Q - (P \cap Q)$. Show that $r(x)$ in $\mathcal{P}_{k+1}$ interpolates $f(x)$ on $P \cup Q$ where

$$r(x) = \frac{(x - \beta)p(x) - (x - \alpha)q(x)}{\alpha - \beta}$$

Solution:

For points $y \in P \cap Q$, we know $p(y) = q(y)$, thus we define $s(y) = p(y) = q(y)$

$$r(y) = \frac{(\alpha - \beta)s(y)}{\alpha - \beta} = s(y) = p(y) = q(y)$$

For point α which is the only point in $P - P \cap Q$

$$r(\alpha) = \frac{(\alpha - \beta)p(\alpha) - 0}{\alpha - \beta} = p(\alpha)$$

For point β which is the only point in $Q - P \cap Q$

$$r(\beta) = \frac{0 - (\beta - \alpha)q(\beta)}{\alpha - \beta} = q(\beta)$$

Hence we covered all the cases for points in $P \cup Q$, and see that $r(x)$ in $\mathcal{P}_{k+1}$ does interpolates $f(x)$ on $P \cup Q$.

University of Washington
Department of Mathematics
MATH 464, Spring 2020

February 23, 2022

Homework 6

Section 5.2.2

Problem 1:
For $f(x) = x^2$ and h arbitrary, calculate $\Delta f(x), \Delta^2 f(x), \Delta^3 f(x), \Delta^4 f(x)$. Repeat this calculation for the function $f(x) = x^3$.

Solution:
Remember that we can find higher-order forward differences using

$$\Delta^k f(x) = \sum_{i=0}^{k} \binom{k}{i} (-1)^i f(x + (k-i)h)$$

Thus the general $\Delta f(x), \Delta^2 f(x)$ is

$$\begin{aligned}
\Delta f(x) &= \binom{1}{0} \cdot 1 \cdot f(x+h) + \binom{1}{1} \cdot (-1) \cdot f(x) \\
&= f(x+h) - f(x) \\
\Delta^2 f(x) &= \binom{2}{0} \cdot 1 \cdot f(x+2h) + \binom{2}{1} \cdot (-1) \cdot f(x+h) + \binom{2}{2} \cdot (1) \cdot f(x) \\
&= f(x+2h) - 2f(x+h) + f(x)
\end{aligned}$$

For particular $f(x) = x^2$

$$\begin{aligned}
\Delta f(x) &= (x+h)^2 - x^2 \\
&= x^2 + 2xh + h^2 - x^2 \\
&= 2xh + h^2 \\
\Delta^2 f(x) &= (x+2h)^2 - 2(x+h)^2 + x^2 \\
&= x^2 + 4xh + 4h^2 - 2x^2 - 4xh - 2h^2 + x^2 \\
&= 2h^2
\end{aligned}$$

For particular $f(x) = x^3$

$$\begin{aligned}
\Delta f(x) &= (x+h)^3 - x^3 \\
&= x^3 + 3x^2h + 3xh^2 + h^3 - x^3 \\
&= 3x^2h + 3xh^2 + h^3 \\
\Delta^2 f(x) &= (x+2h)^3 - 2(x+h)^3 + x^3 \\
&= x^3 + 6x^2h + 12xh^2 + 8h^3 - 2\left(x^3 + 3x^2h + 3xh^2 + h^3\right) + x^3 \\
&= x^3 + 6x^2h + 12xh^2 + 8h^3 - 2x^3 - 6x^2h - 6xh^2 - 2h^3 + x^3 \\
&= 6xh^2 + 6h^3
\end{aligned}$$

Problem 5:
If $p(x)$ interpolates $f(x)$, then we can use $\int_a^b p(x)dx$ as an approximation for $\int_a^b f(x)dx$. Use (5.23) to construct a "numerical integration" formula by choosing $n = 2, x_0 = a$, and $h = (b-a)/2$. That is, verify that

$$\int_a^b p(x)dx = h\int_0^2 p(x_0 + rh)dr$$

and integrate the right-hand side of (5.23). Check your calculations by testing the formula on $f(x) = 1, f(x) = x$, and $f(x) = x^2$ with $a = 0$ and $b = 2$. [The formula should give the correct value for these integrals.] Use the data in Example 5.5 to estimate $\int_3^5 \cos(x)dx$.

Solution:
To verify Let $x = x_0 + rh$.
We then solve for r_0, r_n as the new bounds. We have

$$\begin{aligned}
x_0 &= x_0 + r_0 h, r_0 = 0 \\
x_n &= x_0 + r_n h, r_n = n = 2
\end{aligned}$$

Thus

$$\begin{aligned}
x &= x_0 + rh \\
\frac{dx}{dr} &= h \\
dx &= h\,dr \\
h &= \frac{a-b}{2}\int_a^b p(x)dx = h\int_0^2 p(x_0+h)dr
\end{aligned}$$

By (5.23) we have

$$\begin{aligned}
p(x_0 + rh) &= f(x_0) + \Delta f(x_0)r + \Delta^2 f(x_0)\binom{r}{2} \\
&= f(x_0) + \Delta f(x_0)r + \Delta^2 f(x_0)\frac{r!}{2!(r-2)!} \\
&= f(x_0) + \Delta f(x_0)r + \Delta^2 f(x_0)\frac{r(r-1)}{2}
\end{aligned}$$

Thus integrating the RHS

$$\begin{aligned}
h\int_0^2 p(a+rh)dr &= h\int_0^2 f(a)+\Delta f(a)r+\Delta^2 f(a)\frac{r^2-r}{2}dr \\
&= h\left(f(a)r+\frac{\Delta f(a)}{2}r^2+\frac{\Delta^2 f(a)}{2}\left(\frac{r^3}{3}-\frac{r^2}{2}\right)\right)\Bigg|_0^2 \\
&= h\left(2f(a)+\frac{\Delta f(a)}{2}2^2+\frac{\Delta^2 f(a)}{2}\left(\frac{2^3}{3}-\frac{2^2}{2}\right)\right) \\
&= h\left(2f(a)+2\Delta f(a)+\frac{\Delta^2 f(a)}{2}\left(\frac{2^3}{3}-\frac{2^2}{2}\right)\right) \\
&= h\left(2f(a)+2\Delta f(a)+\frac{\Delta^2 f(a)}{3}\right)
\end{aligned}$$

Then

$$\begin{aligned}
\int_0^2 1dx &= 2 \\
\int_0^2 xdx &= 2 \\
\int_0^2 x^2dx &= \frac{2^3}{3}=\frac{8}{3}
\end{aligned}$$

For $f(x)=1$ the RHS evaluates to

$$\begin{aligned}
h\int_0^2 p(a+rh)dr &= h\left(2f(a)+2\Delta f(a)+\frac{\Delta^2 f(a)}{3}\right) \\
&= h\left(2\cdot 1+2\cdot 0+\frac{0}{2}\left(\frac{2^3}{3}-\frac{2^2}{2}\right)\right) \\
&= 2
\end{aligned}$$

For $f(x)=x$ the RHS evaluates to

$$\begin{aligned}
h\int_0^2 p(a+rh)dr &= h\left(2f(a)+2\Delta f(a)+\frac{\Delta^2 f(a)}{3}\right) \\
&= h\left(2a+2h+\frac{0}{3}\right) \\
&= 1\cdot(2\cdot 0+2\frac{2-0}{2})=2
\end{aligned}$$

For $f(x) = x^2$ the RHS evaluates to

$$\begin{aligned} h\int_0^2 p(a+rh)dr &= h\left(2a^2 + 2(2ah+h^2) + \frac{6ah^2+6h^3}{2}\left(\frac{10}{3}\right)\right) \\ &= h\left(2a^2 + 2(2ah+h^2) + \frac{2h^2}{3}\right) \\ &= 1\cdot\left(0 + 2(0+1) + \frac{2}{3}\right) \\ &= \frac{8}{3} \end{aligned}$$

And thus we conclude that the formula above using $p(x_0 + rh)$ is correct.
For integrating $\int_{0.3}^{0.5}\cos(x)dx$, we find in the divided difference table 5.5

$$\begin{aligned} f(a) &= 0.955336 \\ \Delta f(a) &= -0.034275 \\ \Delta^2 f(a) &= -0.009203 \end{aligned}$$

Thus we have

$$\begin{aligned} h &= \frac{b-a}{2} = \frac{0.5-0.3}{2} = 0.3 \\ h\int_0^2 p(a+rh)dr &= h\left(2f(a) + 2\Delta f(a) + \frac{\Delta^2 f(a)}{3}\right) \\ &= 0.1\cdot(2\cdot 0.955336 + 2\cdot -0.034275 + \frac{-0.009203}{3}) \\ &= 0.183905433333 \\ \int_{0.3}^{0.5}\cos(x)dx &= 0.183905331943 \end{aligned}$$

Problem 8:

Show that $\Delta^m p(x) = 0$ when $p(x) \in \mathcal{P}_n$ and $m \geq n+1$. [Hint: If $p(x) \in \mathcal{P}_n$, show that $\Delta p(x)$, is a polynomial in $\mathcal{P}_{n-1}$; thus conclude that $\Delta^{n-1}p(x)$ is a linear polynomial, $\Delta^n p(x)$ is a constant, and $\Delta^{n+1}p(x) = 0$.]

Lemma 1. If $p(x) \in \mathcal{P}_k$ then $\Delta p(x) \in \mathcal{P}_{k-1}$.

Proof. Consider $p(x) = x^k, p(x) \in \mathcal{P}_k, k \geq 1$. Then $\Delta p(x) = (x+h)^k - x^k$. Using the binomial expansion we see that

$$(x+h)^k = \sum_{j=0}^{k}\binom{k}{j}x^j h^{k-j}$$

There is only one x^k term when $j = k$ in the expansion. Thus, subtracting the x^k gives us at most $k-1$ degree polynomial.
Therefore for $p(x) \in \mathcal{P}_k$, $\Delta p(x) \in \mathcal{P}_{k-1}$. □

Lemma 2. If $p(x)$ is a constant function, $\Delta p(x) = 0$.

Proof. $p(x) = C, \Delta p(x) = p(x+h) - p(x) = C - C = 0$ □

Using Lemma 1, by induction, applying Δ to $p(x)$ n times for $p(x) \in \mathcal{P}_n$ would give us a polynomial of degree 0, hence a constant function. By Lemma 2, applying Δ to $\Delta^n p(x)$, which is a constant function returns 0. Thus for $m = n+1, \Delta^m p(x) = 0$ when $p(x) \in \mathcal{P}_n$. For $m > n+1$, it is easy to see that is simply applying Δ for $m-(n+1)$ times to 0, which is 0. Thus $\Delta^m p(x) = 0$ when $p(x) \in \mathcal{P}_n$ for $m \geq n+1$.

Problem 10:

(pages 222-224 of text) For $n = 1, 2$, and 3, verify that the following version of nested multiplication can be used to evaluate Newton's forward formula for the interpolating polynomial (5.23):
Given any number r, form the numbers $C_n, C_{n-1}, \cdots, C_1$ and C_0 by

$$\begin{aligned} C_n &= \Delta^n f(x_0) \\ C_i &= \Delta^i f(x_0) + (r-i)C_{i+1}/(i+1), \quad i = n-1, n-2, \cdots, 1, 0. \end{aligned}$$

Then $C_0 = p(x_0 + rh)$. Use this iteration to evaluate $p(0.44)$ in Example 5.5. (Note: This form of nested multiplication is valid for all n; but, in general, verification is somewhat difficult exercise.)

Solution:

$$\begin{aligned} h &= 0.1 \\ 0.44 &= 0.3 + 1.4h \\ r &= 1.4 \end{aligned}$$

For $n = 1$:

$$\begin{aligned} C_1 &= \Delta f(x_0) \\ C_0 &= f(x_0) + rC_1 \\ &= f(x_0) + r\Delta f(x_0) \end{aligned}$$

By (5.23)

$$\begin{aligned} p(x_0 + rh) &= f(x_0) + \Delta f(x_0)\binom{r}{1} \\ &= f(x_0) + r\Delta f(x_0) \end{aligned}$$

Evaluating at $p(0.44)$

$$\begin{aligned} C_1 &= -0.034275 \\ C_0 &= 0.955336 + 1.4C_1 \\ &= 0.907351 \end{aligned}$$

For $n = 2$:

$$\begin{aligned}
C_2 &= \Delta^2 f(x_0) \\
C_1 &= \Delta f(x_0) + \frac{(r-1)C_2}{2} \\
C_0 &= f(x_0) + rC_1 \\
&= f(x_0) + r\left(\Delta f(x_0) + \frac{(r-1)C_2}{2}\right) \\
&= f(x_0) + r\left(\Delta f(x_0) + \frac{(r-1)\Delta^2 f(x_0)}{2}\right) \\
&= f(x_0) + r\Delta f(x_0) + \frac{r(r-1)}{2}\Delta^2 f(x_0)
\end{aligned}$$

By (5.23)

$$\begin{aligned}
p(x_0 + rh) &= f(x_0) + \Delta f(x_0)\binom{r}{1} + \Delta^2 f(x_0)\binom{r}{2} \\
&= f(x_0) + r\Delta f(x_0) + \frac{r(r-1)}{2}\Delta^2 f(x_0)
\end{aligned}$$

Evaluating at $p(0.44)$

$$\begin{aligned}
C_2 &= -0.009203 \\
C_1 &= -0.034275 + \frac{(1.4-1)C_2}{2} \\
&= 0.0361156 \\
&= 0.90477416
\end{aligned}$$

For $n = 3$:

$$
\begin{aligned}
C_3 &= \Delta^3 f(x_0) \\
C_2 &= \Delta^2 f(x_0) + \frac{(r-2)C_3}{3} \\
C_1 &= \Delta f(x_0) + \frac{(r-1)C_2}{2} \\
C_0 &= f(x_0) + rC_1 \\
&= f(x_0) + r\left(\Delta f(x_0) + \frac{(r-1)C_2}{2}\right) \\
&= f(x_0) + r\left(\Delta f(x_0) + \frac{(r-1)\left(\Delta^2 f(x_0) + \frac{(r-2)C_3}{3}\right)}{2}\right) \\
&= f(x_0) + r\left(\Delta f(x_0) + \frac{(r-1)\left(\Delta^2 f(x_0) + \frac{(r-2)\Delta^3 f(x_0)}{3}\right)}{2}\right) \\
&= f(x_0) + r\Delta f(x_0) + r\frac{(r-1)\left(\Delta^2 f(x_0) + \frac{(r-2)\Delta^3 f(x_0)}{3}\right)}{2} \\
&= f(x_0) + r\Delta f(x_0) + r\frac{(r-1)\Delta^2 f(x_0) + \frac{(r-1)(r-2)\Delta^3 f(x_0)}{3}}{2} \\
&= f(x_0) + r\Delta f(x_0) + \frac{r(r-1)\Delta^2 f(x_0)}{2} + \frac{r(r-1)(r-2)\Delta^3 f(x_0)}{3 \cdot 2}
\end{aligned}
$$

By (5.23)

$$
\begin{aligned}
p(x_0 + rh) &= f(x_0) + \Delta f(x_0)\binom{r}{1} + \Delta^2 f(x_0)\binom{r}{2} + \Delta^3 f(x_0)\binom{r}{3} \\
&= f(x_0) + r\Delta f(x_0) + \frac{r(r-1)\Delta^2 f(x_0)}{2} + \frac{r(r-1)(r-2)\Delta^3 f(x_0)}{3 \cdot 2}
\end{aligned}
$$

Evaluating at $p(0.44)$

$$
\begin{aligned}
C_3 &= 0.000434 \\
C_2 &= -0.009203 + \frac{(1.4-2)C_3}{3} \\
&= -0.0092898 \\
C_1 &= -0.0034275 + \frac{(1.4-1)C_2}{2} \\
&= -0.03613296 \\
C_0 &= 0.955336 + 1.4C_1 \\
&= 0.904749856
\end{aligned}
$$

Section 5.2.4

Problem 3:

How large should k be if we wish to obtain an interpolation error of 10^{-6} or less throughout $[a, b]$ by interpolating $f(x)$ at the zeros of $\tilde{T}_k(x)$ for

(a) $f(x) = \cos(x), 0.3 \leq x \leq 0.6$

(b) $f(x) = 1/(2 + x), -1 \leq x \leq 1$

(c) $f(x) = \ln(x), 0.1 \leq x \leq 1$

(d) $f(x) = e^{3x}, -1 \leq x \leq 1$

Solution:

(a) By (5.30) and definition of K_k We can find

$$K_k = \max_{0.3 \leq x \leq 0.6} \cos(x) \text{ or } \sin(x)$$

The derivatives of f cycles through $-\sin(x), -\cos(x), \sin(x), \cos(x)$. Plugging in values we see that

$$K_k = \begin{cases} 0.955 & k \text{ is even} \\ 0.565 & k \text{ is odd} \end{cases}$$

Thus by (5.36) we plot the LHS of the inequality

$$\frac{K_k}{2^k(k+1)!} \left(\frac{0.6 - 0.3}{2} \right)^{k+1} \leq 10^{-6}$$

and find that for $k \geq 4$ the inequality is satisfied.

(b) By (5.30) and definition of K_k We can find

$$K_k = \max_{-1 \leq x \leq 1} \frac{(k+1)!}{(2+x)^{k+2}}$$

and easily we can see that the minimum denominator is $(2 + x)^{k+2} = 1$ at $x = -1$. By (5.36) the inequality becomes

$$\begin{aligned} \frac{(k+1)!}{2^k(k+1)!} \left(\frac{(1 - (-1)}{2} \right)^{1+k} &\leq 10^{-6} \\ \frac{1}{2^k} &\leq 10^{-6} \\ 2^k &\geq 10^6 \\ k &\geq \log_2(10^6) \\ k &\geq 19.93 \\ k &\geq 20 \end{aligned}$$

(c) By (5.30) and definition of K_k We can find

$$K_k = \max_{0.1 \le x \le 1} \frac{(k-1)!}{x^{(k+1)}}$$

and easily we can see that the smaller x goes in the denominator, the greater the K_k. Thus K_k is maximized at $x = 0.1$. Also, the numerator continues to grow for each order of the derivative. Thus by (5.36) the inequality we are trying to satisfy becomes

$$\begin{aligned}
\frac{\frac{(k-1)!}{0.1^{(k+1)}}}{2^k(k+1)!}\left(\frac{1-0.1}{2}\right)^{k+1} &\ge 10^{-6} \\
\frac{10^{k+1}}{2^k k(k+1)}\frac{0.9^{k+1}}{2^{k+1}} &\ge 10^{-6} \\
\frac{2.25^{k+1}\cdot 2}{k(k+1)} &\ge 10^{-6}
\end{aligned}$$

The LHS clearly diverges as k increases. Plotting the LHS with k as variable we clearly see that no k gives an output $\le 10^{-6}$.
Thus there is no k we can choose if we wish to obtain an interpolation error of 10^{-6} or less throughout $[a, b]$ by interpolating $f(x)$ at the zeros of $\hat{T}_k(x)$ for $f(x) = \ln(x), 0.1 \le x \le 1$.

(d) By (5.30) and definition of K_k We can find

$$K_k = \max_{-1 \le x \le 1} 3^{k+1}e^x$$

Thus $K_k = 3^{k+1}e^1$ as $x = 1$ maximizes K_k for the interval. Thus by (5.36) we have the inequality

$$\frac{3^{k+1}e}{2^k(k+1)!}\left(\frac{1-(-1)}{2}\right)^{k+1} \le 10^{-6}$$

plotting the LHS with different k we see that for $k \ge 13$ the inequality is satisfied.

Problem 5:

$$W(x) = (x - x_0)(x - x_1)\cdots(x - x_n)$$

Since $x_0, x_1, \cdots, x_n$ is equally spaced with n sub-interval h over the interval of $[-1, 1]$, we could see the endpoint x_0 and x_n as the ordinal position $N = n/2$ away from the midpoint $x_{n/2}$. Thus $x_0 = -1 = -Nh, x_1 = 1 = Nh, x_{n/2} = 0$.
Thus

$$\begin{aligned}
W(x) &= (x - x_0)(x - x_1)\cdots(x - x_n) \\
&= (x - (-Nh))(x - (-Nh + h))(x - (-Nh + 2h))\cdots \\
&(x - (-Nh + (N-1)h))(x - (-Nh + Nh))(x - (-Nh + (N+1)h))\cdots \\
&(x - (-Nh + (N + (N-2))h))(x - (-Nh + (N + (N-1))h))(x - (-Nh + (N+N)h)) \\
&= (x + Nh)(x + (N-1)h)\cdots(x + h)x(x - h)\cdots(x - (N-1)h)(x - Nh)
\end{aligned}$$

We are now considering $x_{n-1} < x < x_n$ so that $N-1 < r < N$ for $x \in [-1,1]$. We should find the maximum as $x \in [x_i, x_{i+1}]$ for $i = 0$ and $i = n$. At the endpoint intervals, each interpolating point follows the pattern below:

$$|x - x_{i+2}| \leq 2h, |x - x_{i+3}| \leq 3h, |x - x_{i+4}| \leq 4h, \cdots |x - x_{i+n}| \leq nh$$
$$|x - x_{i-1}| \leq 2h, |x - x_{i-2}| \leq 3h, |x - x_{i-3}| \leq 4h, \cdots |x - x_{i-(n-1)}| \leq nh$$

Thus, there are $(n-1)$ intervals in total. Thus there are

$$\underbrace{(2h)(3h)(4h)(5h)\cdots(nh)}_{n-1 \text{ intervals}} = n!h^{n-1}$$

The lower bound follows suits, as we imagine x being on the nearest end to the other intervals:

$$|x - x_{i+2}| \leq h, |x - x_{i+3}| \leq 2h, |x - x_{i+4}| \leq 3h, \cdots |x - x_{i+n}| \leq (n-1)h$$
$$|x - x_{i-1}| \leq h, |x - x_{i-2}| \leq 2h, |x - x_{i-3}| \leq 3h, \cdots |x - x_{i-(n-1)}(n-1)h$$

Similarly we get the lower bound

$$\underbrace{(h)(2h)(3h)(4h)\cdots((n-1)h)}_{n-1 \text{ intervals}} = (n-1)!h^{n-1}$$

Thus we conclude

$$(n-1)!h^{n-1}|(x - x_{n-1})(x - x_n)| \leq W(x) \leq n!h^{n-1}|(x - x_{n-1})(x - x_n)|.$$

Problem 6:

For a given fixed $x \in [x_i, x_{i+1}]$, it is easy to see that for the factor $|(x - x_{n-1})(x - x_n)|$ is maximized at the midpoint of the interval $x = \frac{x_i + x_{i+1}}{2}$ by calculus. Since it is at the midpoint, $(x - x_{n-1}) = (x - x_n) = \frac{h}{2}$. Thus the maximum value of the factor $|(x - x_{n-1})(x - x_n)|$. Substituting this into the inequality in problem 5 with $h = \frac{2}{n}$, we have

$$n!h^{n-1}|(x - x_{n-1})(x - x_n)| = \frac{1}{4}n!h^{n+1} = \frac{1}{4}n!\left(\frac{2}{n}\right)^{n+1} = \frac{n!2^{n-1}}{n^{n+1}}$$
$$(n-1)!h^{n-1}|(x - x_{n-1})(x - x_n)| = \frac{1}{4}(n-1)!h^{n+1} = \frac{1}{4}(n-1)!\left(\frac{2}{n}\right)^{n+1} = \frac{(n-1)!2^{n-1}}{n^{n+1}}$$

Thus we can conclude that

$$\frac{(n-1)!2^{n-1}}{n^{n+1}} \leq |W(x)| \leq \frac{n!2^{n-1}}{n^{n+1}}$$

Problem 10:

$$W(x) = (x+nh)(x+(n-2)h)\cdots(x+h)(x-h)\cdots(x-(n-2)h)(x-nh)$$
$$W(x) = (x+nh)(x-nh)(x+(n-2)h)(x-(n-2)h)\cdots(x+h)(x-h)$$
$$W(x) = (x^2-n^2h^2)(x^2-(n-2)^2h^2)(x^2-(n-4)^2h^2)\cdots(x^2-h^2)$$

Since n is odd, $x_0, x_1, \cdots, x_n$ has even number of terms. It is easy to see that $W(x)$ is an even function.
Thus, $W'(x)$ is a odd function. Thus $W'(x)$ and $W'(-x)$ has opposite sign. Thus only for $x=0$ do we have a derivative of 0. For an even function the maximum or minimum occurs at a point with derivative of 0.
We can further the argument as there are $\frac{n+1}{2}$ terms, which means there are still even number of terms of the gathered terms.
Thus we can switch the order of x^2 and $(n-i)^2h^2$. Thus it is clear that each term is maximized when $x=0$ since we subtract 0.
For $f(x) = \frac{1}{x+2}$ with $x \in (-h, h)$

$$K_k = \max_{-h\le x\le h} \frac{(k+1)!}{(2+x)^{k+2}}$$

and the upper bound is reached at $x=-h$. Thus the we have

$$\begin{aligned}
|e(x)| &\le \frac{\frac{(n+1)!}{(2-h)^{n+2}}}{2^n(n+1)!}W(0)\\
&\le \frac{\frac{(n+1)!}{(2-h)^{n+2}}}{2^n(n+1)!}W(0)\\
&\le \frac{1}{2^n\,(2-h)^{n+2}}W(0)\\
W(0) &= (0-n^2h^2)(0-(n-2)^2h^2)(0-(n-4)^2h^2)\cdots(0-h^2)\\
&= \underbrace{(n^2h^2)((n-2)^2h^2)((n-4)^2h^2)\cdots(h^2)}_{\frac{n+1}{2}\text{ terms}}\\
&= n^2(n-2)^2(n-4)^2\cdots 2^2h^{n+1}\\
&= \frac{(n+1)!}{2^{\frac{n+1}{2}}\left(\frac{n+1}{2}\right)!}h^{n+1}\\
|e(x)| &\le \frac{1}{2^n\,(2-h)^{n+2}}\frac{(n+1)!}{2^{\frac{n+1}{2}}\left(\frac{n+1}{2}\right)!}h^{n+1}
\end{aligned}$$

Section 5.2.6

Problem 1:

(a)

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 4 & 12 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 2 \\ 1 \end{bmatrix}$$

$$\begin{aligned} a_0 &= 1 \\ a_1 &= 1 \\ a_1 + 4a_2 + 12a_3 &= 2 \\ 1 + 4a_2 + 12a_3 &= 2 \\ 4a_2 + 12a_3 &= 1 \\ 1 + 1 + a_2 + a_3 &= 1 \\ a_2 + a_3 &= -1 \\ 8a_2 &= 5 \\ a_3 &= \frac{5}{8} \\ a_2 &= -\frac{13}{8} \end{aligned}$$

There is a unique solution $a_0 = 1, a_1 = 1, a_2 = -\frac{13}{8}, a_4 = \frac{5}{8}$.

(b)

$$\begin{bmatrix} 1 & -1 & 1 & -1 \\ 0 & 1 & -2 & 3 \\ 0 & 1 & 2 & 3 \\ 1 & 2 & 4 & 8 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 2 \\ 1 \end{bmatrix}$$

We can do row reduction on this matrix:

$$\left[\begin{array}{cccc|c} 1 & -1 & 1 & -1 & 1 \\ 0 & 1 & -2 & 3 & 1 \\ 0 & 1 & 2 & 3 & 2 \\ 0 & 3 & 3 & 9 & 0 \end{array}\right]$$

$$\left[\begin{array}{cccc|c} 1 & -1 & 1 & -1 & 1 \\ 0 & 1 & -2 & 3 & 1 \\ 0 & 1 & 2 & 3 & 2 \\ 0 & 1 & 1 & 3 & 0 \end{array}\right]$$

$$\left[\begin{array}{cccc|c} 1 & -1 & 1 & -1 & 1 \\ 0 & 1 & -2 & 3 & 1 \\ 0 & 0 & 4 & 0 & 1 \\ 0 & 1 & 1 & 3 & 0 \end{array}\right]$$

$$\left[\begin{array}{cccc|c} 1 & -1 & 1 & -1 & 1 \\ 0 & 1 & -2 & 3 & 1 \\ 0 & 0 & 4 & 0 & 1 \\ 0 & 0 & 3 & 0 & -1 \end{array}\right]$$

$$a_2 = \frac{1}{4} \neq \frac{-1}{3} = a_2$$

There is no solution to this system of linear equations.

(c)

$$\begin{bmatrix} 1 & -1 & 1 & -1 \\ 0 & 1 & -2 & 3 \\ 0 & 1 & 2 & 3 \\ 1 & 2 & 4 & 8 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} 1 \\ -6 \\ 2 \\ 1 \end{bmatrix}$$

We can do row reduction on this matrix:

$$\left[\begin{array}{cccc|c} 1 & -1 & 1 & -1 & 1 \\ 0 & 1 & -2 & 3 & -6 \\ 0 & 1 & 2 & 3 & 2 \\ 0 & 3 & 3 & 9 & 0 \end{array}\right]$$

$$\left[\begin{array}{cccc|c} 1 & -1 & 1 & -1 & 1 \\ 0 & 1 & -2 & 3 & -6 \\ 0 & 1 & 2 & 3 & 2 \\ 0 & 1 & 1 & 3 & 0 \end{array}\right]$$

$$\left[\begin{array}{cccc|c} 1 & -1 & 1 & -1 & 1 \\ 0 & 1 & -2 & 3 & -6 \\ 0 & 0 & 4 & 0 & 8 \\ 0 & 1 & 1 & 3 & 0 \end{array}\right]$$

$$\left[\begin{array}{cccc|c} 1 & -1 & 1 & -1 & 1 \\ 0 & 1 & -2 & 3 & -6 \\ 0 & 0 & 4 & 0 & 8 \\ 0 & 0 & 3 & 0 & 6 \end{array}\right]$$

$$\left[\begin{array}{cccc|c} 1 & -1 & 1 & -1 & 1 \\ 0 & 1 & -2 & 3 & -6 \\ 0 & 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 0 & 2 \end{array}\right]$$

$$\left[\begin{array}{cccc|c} 1 & -1 & 1 & -1 & 1 \\ 0 & 1 & -2 & 3 & -6 \\ 0 & 0 & 1 & 0 & 2 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

$$\left[\begin{array}{cccc|c} 1 & -1 & 1 & -1 & 1 \\ 0 & 1 & 0 & 3 & -2 \\ 0 & 0 & 1 & 0 & 2 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

$$\left[\begin{array}{cccc|c} 1 & 0 & 1 & 2 & -1 \\ 0 & 1 & 0 & 3 & -2 \\ 0 & 0 & 1 & 0 & 2 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

$$\left[\begin{array}{cccc|c} 1 & 0 & 0 & 2 & -3 \\ 0 & 1 & 0 & 3 & -2 \\ 0 & 0 & 1 & 0 & 2 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

$$a_0 = -a, a_1 = a, a_2 = 2, a_3 = -a$$

There are infinitely many solutions since a_3 is a free variable.

Problem 3:

$$\begin{bmatrix} 1 & -2 & 4 & 8 \\ 0 & 1 & -4 & 12 \\ 1 & 0 & 0 & 0 \\ 1 & 2 & 4 & 8 \\ 0 & 1 & 4 & 12 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \\ a_4 \end{bmatrix} = \begin{bmatrix} 2 \\ -1 \\ 0 \\ 2 \\ 1 \end{bmatrix}$$

We can do row reduction on this matrix:

$$\Rightarrow \left[\begin{array}{ccccc|c} 1 & -2 & 4 & -8 & 16 & 2 \\ 0 & 1 & -4 & 12 & -32 & -1 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 2 & 4 & 8 & 16 & 2 \\ 0 & 1 & 4 & 12 & 32 & 1 \end{array}\right] \Rightarrow \left[\begin{array}{ccccc|c} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & -2 & 4 & -8 & 16 & 2 \\ 0 & 2 & 4 & 8 & 16 & 2 \\ 0 & 1 & -4 & 12 & -32 & -1 \\ 0 & 1 & 4 & 12 & 32 & 1 \end{array}\right]$$

$$\Rightarrow \left[\begin{array}{ccccc|c} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 4 & 8 & 16 & 2 \\ 0 & 0 & 8 & 0 & 32 & 4 \\ 0 & 1 & -4 & 12 & -32 & -1 \\ 0 & 2 & 0 & 24 & 0 & 0 \end{array}\right] \Rightarrow \left[\begin{array}{ccccc|c} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 4 & 8 & 16 & 2 \\ 0 & 0 & 8 & 0 & 32 & 4 \\ 0 & 1 & -4 & 12 & -32 & -1 \\ 0 & 2 & 0 & 24 & 0 & 0 \end{array}\right]$$

$$\Rightarrow \left[\begin{array}{ccccc|c} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 4 & 8 & 16 & 2 \\ 0 & 0 & 8 & 0 & 32 & 4 \\ 0 & 1 & -4 & 12 & -32 & -1 \\ 0 & 2 & 0 & 24 & 0 & 0 \end{array}\right] \Rightarrow \left[\begin{array}{ccccc|c} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 2 & 4 & 8 & 1 \\ 0 & 0 & 1 & 0 & 4 & \frac{1}{2} \\ 0 & 1 & -4 & 12 & -32 & -1 \\ 0 & 1 & 0 & 12 & 0 & 0 \end{array}\right]$$

$$\Rightarrow \left[\begin{array}{ccccc|c} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 2 & 4 & 8 & 1 \\ 0 & 0 & 1 & 0 & 4 & \frac{1}{2} \\ 0 & 0 & 6 & -8 & 40 & 2 \\ 0 & 0 & 2 & -8 & 8 & 1 \end{array}\right] \Rightarrow \left[\begin{array}{ccccc|c} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 2 & 4 & 8 & 1 \\ 0 & 0 & 1 & 0 & 4 & \frac{1}{2} \\ 0 & 0 & 0 & 8 & -16 & 1 \\ 0 & 0 & 0 & -8 & 0 & 0 \end{array}\right]$$

$$\Rightarrow \left[\begin{array}{ccccc|c} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & \frac{3}{4} \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -\frac{1}{16} \end{array}\right]$$

Thus the polynomial is $p*(x) = \frac{3}{4}x^2 - \frac{1}{16}x^4$.The derivative is $p*'(x) = \frac{3}{2}x - \frac{1}{4}x^3$.
Comparing to the cubic spline and the interpolating polynomial we have

x	$f(x)$	$S_3(x)$	$S_3'(x)$	$p(x)$	$p'(x)$	$p*(x)$	$p*'(x)$
1.0	1.0	1.000	1.286	1.000	1.667	1.000	0.688
1.2	1.2	1.241	1.131	1.334	1.648	1.334	1.408
1.4	1.4	1.455	1.011	1.646	1.437	1.646	1.354
1.6	1.6	1.648	0.926	1.894	2.030	1.894	1.142
1.8	1.8	1.827	0.874	2.030	0.312	2.030	0.954
2.0	2.0	2.000	0.857	2.000	−0.667	2.000	1

and we see that this polynomial is generally better than the interpolating polynomial but not as good as the cubic spline.